



Stochastic modeling, analysis, and simulation of the COVID-19 pandemic with explicit behavioral changes in Bogotá: A case study



David Niño-Torres^a, Andrés Ríos-Gutiérrez^a, Viswanathan Arunachalam^{a,*},
Comfort Ohajunwa^b, Padmanabhan Seshaiyer^b

^a Department of Statistics, Universidad Nacional de Colombia, Bogotá, Colombia

^b George Mason University, Fairfax, VA, USA

ARTICLE INFO

Article history:

Received 20 July 2021

Received in revised form 24 December 2021

Accepted 24 December 2021

Available online 31 December 2021

Handling Editor: Dr HE DAIHAI HE

Keywords:

Epidemic spread

SARS-CoV-2

Compartmental modeling

Stochastic perturbations

Parameter estimation

ABSTRACT

In this paper, a stochastic epidemiological model is presented as an extension of a compartmental SEIR model with random perturbations to analyze the dynamics of the COVID-19 pandemic in the city of Bogotá D.C., Colombia. This model incorporates the spread of COVID-19 impacted by social behaviors in the population and allows for projecting the number of infected, recovered, and deceased individuals considering the mitigation measures, namely confinement and partial relaxed restrictions. Also, the role of randomness using the concept of Brownian motion is emphasized to explain the behavior of the population. Computational experiments for the stochastic model with random perturbations were performed, and the model is validated through numerical simulations for actual data from Bogotá D.C.

© 2021 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The coronavirus disease (COVID-19) that is believed to have emerged in Wuhan, China, in late 2019 spread to Colombia on March 6, 2020, when the country had reported its first positive case (Benítezet al., 2020; Ministerio de Salud y Protección Social, 2020). The same month, the World Health Organization had declared the disease as a pandemic (World Health Organization, 2020). These developments have prompted much research dedicated to understanding the nature of the spread of the disease (Sohrabiet al., 2020; Wanget al., 2020; Wu et al., 2020). Specifically, various mathematical models have been widely employed to predict the future spread of the disease as well as to explore the impact of certain interventions on the spread (see (Amouch & Karim, 2021; Linet al., 2020; Musae et al., 2021)).

Most mathematical models describing the spread of the disease employ classical compartments, such as the Susceptible-Exposed-Infected-Recovered (SEIR) structure described as an ordinary differential equation system (Brauer and Castillo-Chavez, 2012). Some researchers have adapted these models to include several relevant factors, including the role of containment strategies (Maier and Brockmann, 2020), mobility patterns (Rădulescu et al., 2020), social distancing

* Corresponding author.

E-mail address: varunachalam@unal.edu.co (V. Arunachalam).

Peer review under responsibility of KeAi Communications Co., Ltd.

(Leunget al., 2018; Premet al., 2020), and network intervention strategies to patterns of plateau duration, intensity and duration of social distancing measures (Komarova et al., 2021).

However, the pandemic has prompted lock-downs, widespread closures, and calls to social distance and practice basic hygiene, which has encouraged many people to take caution to limit the spread (Alberto and GongGiordano, 2021; Orabyet al., 2021). To account for the impact of such non-pharmaceutical interventions and the social behavior of the population in response, recently a novel mathematical framework was developed (Ohajunwa et al., 2020). Specifically, three separate models including a baseline model, an explicit intervention model, and an implicit intervention model were created. Of particular relevance to this paper is the explicit intervention model, which accounted for the effect of lock-down through the addition of a Confined compartment to an extended SEIR model. Based on Dirac delta functions, a portion of susceptible individuals was modeled to enter confinement during lock-down and return to become susceptible when lock-down ended. While this model was able to incorporate social behavior through the transmission parameters, this model did not include the influence of stochasticity. Given that many individuals within the population are assumed to react irrationally and unpredictably in response to the pandemic, there is a need to consider stochastic models to capture these effects. Stochastic perturbation models can be used to study the effects of these reactions by introducing environmental noise on the transmission parameter (see (Cai et al., 2019; Grayet al., 2011; Ríos-Gutiérrez et al., 2021)).

Colombia, like many other countries, has implemented a series of emergency measures in response to the coronavirus outbreak. The government quickly responded by policy implementation by introducing strict lock-downs, PCR testing capacity, contact tracing, and augmenting ICU capacity in the hospitals. While Colombia's management of COVID-19 may be considered as a success, the country was not been able to flatten the curve for more than 100 days at the beginning of the pandemic (see (Cárdenas & Martínez, 2020; De la Hoz-Restrepo et al., 2020)). Detailed daily reports and other information can be found at (Instituto Nacional de Salud de Colombia; Ministerio de Salud y Protección Social, 2020).

Thus, the goal of this paper is to develop a novel mathematical framework that extends the previous model (Ohajunwa et al., 2020) to not only account for the effects of social behavior, specifically during lock-down conditions, but also the role of stochasticity within the population. Furthermore, we also apply this model to real data from the Colombian city of Bogotá through computational techniques. Specifically, we use data from the National Institute of Health of Colombia (see (Instituto Nacional de Salud de Colombia)) to derive parameters for our simulations and analysis. This model will project the number of infected, recovered and deceased individuals taking into account not only mitigation measures, specifically confinement and partial opening, but also the role of randomness in the behavior of the population in the pandemic.

This paper will be structured as follows. In section 2, we introduce the stochastic ordinary differential equation system, including the definition of the variables and parameters. The section also describes the numerical discretization for this system along with calibration of the model. In section 3, we validate the model by performing numerical simulations against real data from Bogotá. Finally, in section 4, we discuss and draw conclusions from our work.

2. A stochastic model: equations and methods

To address the disease dynamics of the COVID-19 pandemic in the city of Bogotá DC, we propose a stochastic compartmental disease transmission model based on (Ohajunwa et al., 2020) and adapted according to a structure of stochastic differential equations. The phases or compartments are based on the scientific literature and the different decrees adopted by the various governmental and health institutions in charge.

The flow diagram of the proposed mathematical model is shown in Fig. 1. Specifically, individuals transition from one compartment to another according to specific rates and are influenced by noise, which simulates the random dynamics of the epidemic. The model assumes that the total population is homogeneous and constant in size N . Immigration and emigration are not taken into account; that is, the population is closed.

For this work, we introduce the following the sub-populations: Susceptible (S), Confined (C), Exposed (E), Asymptomatic (I_A), Symptomatic (I_S), Quarantined (Q), Hospitalized (H), Recovered (R), and Dead (D). Susceptible individuals consist of individuals who are not infected with COVID-19 and are not isolated from the population. Confined individuals are individuals who were previously susceptible but by their own or regulatory decision are temporarily isolated from the population in order to protect themselves from infection. Exposed individuals are individuals who are no longer susceptible because they have come into contact with asymptomatic or symptomatic infected individuals and are in the incubation period of disease progression. Asymptomatic individuals are infected individuals who do not exhibit any symptoms of COVID-19, while symptomatic individuals are infected individuals who do have symptoms. Hospitalized individuals are individuals who have been symptomatic and enter hospitalization with severe COVID-19 symptoms. Quarantined individuals are symptomatic individuals who have been isolated. Note that in this model, we suppose that symptomatic population immediately go to quarantine and does not include any death of this population. However, with more available data, we could consider this case in the future. Recovered individuals consist of individuals who were previously infectious and has now survived COVID-19. Deceased individuals are individuals who were infectious but did not survive COVID-19. Given that, we assume a closed population size, $N = S + C + E + I_A + I_S + H + Q + R + D$.

In addition, we define the parameters used for the model in Table 1. The rates at which susceptible and confined individuals enter and leave confinement depend on the Dirac delta functions $\varphi(t)$ and $\psi(t)$. Susceptible individuals become exposed upon contact with the infected population at the transmission rate β . Exposed individuals become infected at the incubation rate, κ , such that a fraction p becomes symptomatic infectious and the remaining $(1 - p)$ becomes asymptomatic

infectious. Asymptomatic individuals either recover or develop symptoms at a rate of ω , with a fraction ν recovering and a fraction $(1 - \nu)$ becoming symptomatic. Symptomatic individuals quarantine at a rate of ξ . At a rate of γ , a fraction q of quarantined individuals are hospitalized and a fraction $(1 - q)$ recover. Lastly, a proportion x of hospitalized individuals die at a rate of α^D , while a proportion $(1 - x)$ of hospitalized individuals recover at a rate of α^R .

Next, we propose the following system of stochastic differential equations for the *modified Stochastic SEIR model with random perturbations* (see (Ríos-Gutiérrez et al., 2021)). Note that this model will include random perturbations of the white noise terms satisfying the Brownian motion governed by system of stochastic differential equations. Let $\{S(t), E(t), I^A(t), I^S(t), Q(t), H(t), R(t), D(t), C(t)\}_{t \geq 0}$ be an Itô process given by the following system of differential equations:

$$\begin{aligned}
 dS(t) &= \left[-\phi S(t) + \psi C(t) - S(t)\tilde{\beta} \left(\frac{I^A(t) + I^S(t)}{N} \right) \right] dt - \frac{\sigma_1 S(t) I^A(t)}{N} dB_1(t) - \frac{\sigma_2 S(t) I^S(t)}{N} dB_2(t) \\
 \frac{dE(t)}{dt} &= \tilde{\beta} S(t) \left(\frac{I^A(t) + I^S(t)}{N} \right) - \kappa E(t) \\
 dt^A(t) &= \left((1-p)\kappa E(t) - \omega I^A(t) \right) dt + \frac{\sigma_1 S(t) I^A(t)}{N} dB_1(t) \\
 dt^S(t) &= \left(p\kappa E(t) + (1-\nu)\omega I^A(t) - \xi I^S(t) \right) dt + \frac{\sigma_2 S(t) I^S(t)}{N} dB_2(t) \\
 \frac{dQ(t)}{dt} &= \xi I^S(t) - \gamma Q(t) \\
 \frac{dH(t)}{dt} &= q\gamma Q(t) - ((1-x)\alpha_R + x\alpha_D)H(t) \\
 \frac{dR(t)}{dt} &= \nu\omega I^A(t) + (1-q)\gamma Q(t) + (1-x)\alpha_R H(t) \\
 \frac{dD(t)}{dt} &= x\alpha_D H(t) \\
 \frac{dC(t)}{dt} &= \phi S(t) - \psi C(t),
 \end{aligned} \tag{1}$$

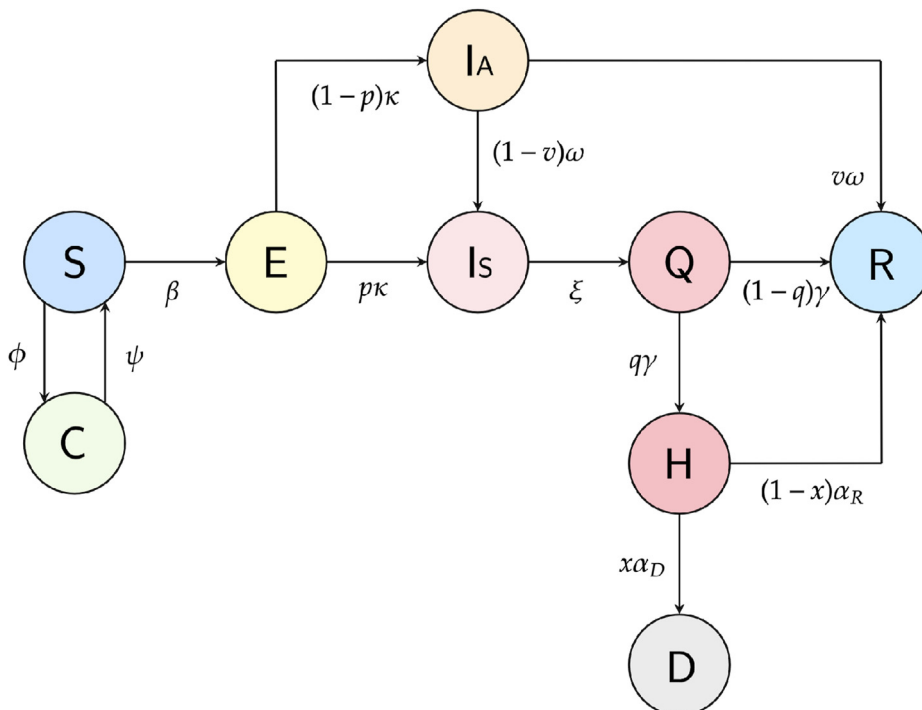


Fig. 1. A flow diagram of the proposed model dynamics.

Table 1
Definition of Parameters in the model.

Parameter	Definition
β	Transmission Rate
φ	Rate at which susceptible individuals enter confinement as a function of time
ψ	Rate at which confined individuals reenter the susceptible sub-population as a function of time
κ	Incubation rate
ω	Rate at which asymptomatic individuals become symptomatic or recover
ξ	Rate at which symptomatic individuals enter quarantine
γ	Rate at which quarantined individuals become hospitalized or recover
α_R	Recovery rate for hospitalized individuals
α_D	Death rate for hospitalized individuals
p	Proportion of exposed individuals that become symptomatic
q	Proportion of quarantined individuals that become hospitalized
ν	Proportion of asymptomatic individuals that recover
x	Proportion of hospitalized individuals that die
l	Proportion of susceptible individuals that enter confinement at $t = t_l$
m	Proportion of confined individuals that reenter the susceptible population at $t = t_m$
t_l	Lock-down release time (in days from $t = 0$)
t_m	Lock-down time (in days from $t = 0$)

where $\{B_1(t)\}_{t \geq 0}$ and $\{B_2(t)\}_{t \geq 0}$ are two independent standard Brownian motions, and considered as generalized white noise functionals, and the constants σ_1 and σ_2 are the intensities of the noises. We note that $\{B_1(t)\}_{t \geq 0}$ and $\{B_2(t)\}_{t \geq 0}$ are defined on the complete probability space $(\Omega, \mathfrak{F}, \{\mathfrak{F}\}_{t \geq 0}, \mathbb{P})$ satisfying the usual conditions. The functions $\varphi = \varphi(t) = l\delta(t - t_l)$ and $\psi = \psi(t) = m\delta(t - t_m)$, such that:

$$\delta(t - t_x) = \begin{cases} 1 & \text{if } t - t_x = 0 \\ 0 & \text{if } t - t_x \neq 0 \end{cases} \tag{2}$$

Since the transmission rate changes according to the dynamics of the confinement or lock, then the following is proposed:

$$\tilde{\beta}(t - t_l) = \begin{cases} (1 + \epsilon)\beta & \text{if } t - t_l > 0 \\ \beta & \text{if } t - t_l \leq 0. \end{cases} \tag{3}$$

2.1. Basic reproduction number

We now briefly discuss the basic reproduction number R_0 , which is the expected number of secondary cases produced by a single infection in a completely susceptible population. Using the method for the epidemic models with random perturbations (Ríos-Gutiérrez et al., 2021), we define the basic reproduction number as follows:

$$R_0 := \int_0^\infty b(a)F(a)da, \tag{4}$$

where $b(a)$ is the average number of newly infected individuals (in a completely susceptible population) by an infected individual that is infectious between $t = 0$ and $t = a$. $F(a)$ is the probability that a newly infected individual will continue infecting others during the time interval between 0 and a . This is also called the underlying survival probability (or survival function).

In our case, the infectivity rate when the population is completely susceptible is given by

$$\tilde{\beta}_1 := \frac{\beta}{N} + \frac{\sigma_1}{N}B_1(t)$$

if the first infected individual is asymptomatic, or

$$\tilde{\beta}_2 = \frac{\beta}{N} + \frac{\sigma_2}{N}B_2(t)$$

if the first infected individual is symptomatic. When the first infected individual infects a completely susceptible population with N individuals, then the expected number of asymptomatic individuals who were infected by a symptomatic individual is

$$b_1(a) = N\tilde{\beta}_2(1 - p) = (\beta + \sigma_2 E(B_2(+\infty)))(1 - p),$$

the expected number of symptomatic individuals who were infected by a symptomatic individual is

$$b_2(a) = N\tilde{\beta}_2 p = \beta + \sigma_2 E(B_2(+\infty))p,$$

and the expected number of asymptomatic individuals who were infected by an asymptomatic individual and have not recovered from the illness, that is, became symptomatic individuals is

$$b_3(a) = N\tilde{\beta}_1(1 - v)(1 - p) = (\beta + \sigma_1 E(B_1(+\infty)))(1 - v)(1 - p),$$

and the expected number of symptomatic individuals who were infected by an asymptomatic individual is

$$b_4(a) = N\tilde{\beta}_1 p = \beta + \sigma_1 E(B_1(+\infty))p.$$

Note that we wrote $E(B_i(+\infty)) = \lim_{t \rightarrow +\infty} B_i(t) = 1, 2$ due to the need to establish how many people are infected by an infectious individual, which we can interpret as the time of infecting from the beginning to the end of time.

In this way,

$$(\beta + \sigma_2 B_2(a))p + (\beta + \sigma_2 B_2(a))(1 - p) + (\beta + \sigma_1 B_1(a))(1 - v)(1 - p) + (\beta + \sigma_1 B_1(a))p$$

is the number of infected individuals by an infectious individual for $a \rightarrow +\infty$. We have the survival functions, $F_1(a) = e^{-\xi a}$ and $F_2(a) = e^{-\omega a}$, for when (1) a symptomatic individual infected by a symptomatic individual remains symptomatic on $[0, a]$, (2) an asymptomatic individual infected by a symptomatic individual remains asymptomatic on $[0, a]$ and (3) an asymptomatic individual infected by an asymptomatic individual remains infectious on $[0, a]$, and (4) a symptomatic individual infected by an asymptomatic individual remains symptomatic on $[0, a]$; respectively. We obtain the following equation

$$R_0^{cov} := \int_0^{+\infty} b(a)F(a)da = \int_0^{+\infty} \beta p F_1(a)da + \int_0^{+\infty} \beta(1 - p)F_2(a)da + \int_0^{+\infty} \beta(1 - v)(1 - p)F_1(a)da + \int_0^{+\infty} \beta p F_1(a)da$$

Therefore, in the following equation we introduce $b_e(a)$ and $F_e(a)$ which are functions corresponding to $b(a)$ and $F(a)$, respectively, for the stochastic system (1). We then have,

$$\begin{aligned} \int_0^{+\infty} b_e(a)F_e(a)da &= 2\frac{\beta p}{\xi} + \frac{\beta(1 - p)}{\omega} + \frac{\beta(1 - v)(1 - p)}{\xi} + \int_0^{+\infty} \sigma_2 B_2(a)(pe^{-\xi a} + (1 - p)e^{-\omega a})da \\ &+ \int_0^{+\infty} \sigma_1((1 - v)(1 - p) + p)B_1(a)e^{-\xi a}da. \end{aligned} \tag{5}$$

The integral $\int_0^{+\infty} b_e(a)F_e(a)da$ is the mean of the random variable that describes the basic reproduction number for the proposed stochastic model. Based on (Ríos-Gutiérrez et al., 2021), we have that

$$R_1 := \int_0^{+\infty} \sigma_1((1 - v)(1 - p) + p)B_1(a)e^{-\xi a}da \sim \mathbf{N}\left(0, \frac{\sigma_1^2((1 - v)^2(1 - p) + p)^2}{2\xi^3}\right) \tag{6}$$

On the other hand, using integration-by-parts rule (Chung and Williams, 2014), we have that

$$\lim_{l \rightarrow +\infty} B_2(l)(\omega pe^{-\xi l} + \xi(1 - p)e^{-\omega l}) = B_2(0)(p\omega + (1 - p)\xi) - I_1 - I_2 + I_3$$

where,

$$I_1 = \int_0^{+\infty} \omega p \xi B_2(a) e^{-\xi a} da I_2 = \int_0^{+\infty} \xi (1-p) \omega B_2(a) e^{-\omega a} da I_3 = \int_0^{+\infty} (\omega p e^{-\xi a} + \xi (1-p) e^{-\omega a}) dB_2(a),$$

We compute that $\lim_{l \rightarrow +\infty} B_2(l)(\omega p e^{-\xi l} + \xi (1-p) e^{-\omega l}) = 0$ a.s. In addition, $B_2(0) = 0$ a.s. Now using (Klebaner, 2012)[p. 393], we obtain

$$R_2 := \int_0^{+\infty} \sigma_2 B_2(a) (p e^{-\xi a} + (1-p) e^{-\omega a}) da = \int_0^{+\infty} \sigma_2 (\omega p e^{-\xi a} + \xi (1-p) e^{-\omega a}) dB_2(a) \sim \mathbf{N}\left(0, \frac{\sigma_2^2 (1-p)^2}{2\omega^3} + \frac{2\sigma_2^2 p (1-p)}{(\xi + \omega)\xi\omega}\right)$$

since we assumed $\{B_1(t)\}_{t \geq 0}$ and $\{B_2(t)\}_{t \geq 0}$ are two Brownian independent motions, in consequence, we get

$$R_{0,v}^{Cov} = \int_0^{+\infty} b_e(a) F_e(a) da = R_0^{Cov} + R_2 + R_1 \sim \mathbf{N}\left(R_0^{Cov}, Var(R_1) + Var(R_2)\right). \tag{7}$$

Thus, we obtain the distribution of the basic reproduction number for the proposed model.

2.2. Numerical discretization

We now briefly explain the numerical approximation method used in the paper. The Euler-Maruyama method described in (Kloeden and Platen, 1992) and introduced by the Japanese mathematician G. Maruyama (Maruyama, 1955) as an extension of the Euler method, is a numerical integration technique for obtaining approximate solutions for a system of stochastic differential equations from a given initial value $X_0 = x_0$. Let $0 = t_0 < t_1 < \dots < t_{k-1} < t_k = T$ be a partition of the interval $[0, T]$, where the length of each subinterval is $\Delta t = t_{i+1} - t_i = T/k$, which implies that $t_{i+1} = t_i + \Delta t = i\Delta t$ and $\Delta B_i = \Delta B(t_i) = B(t_i + \Delta t) - B(t_i)$. For each stochastic process trajectory, the value of $X_{t_{i+1}}$ is approximated using only the value of the previous time step, X_{t_i} . Then, to find the trajectories or approximate solutions of a stochastic differential equation by the Euler-Maruyama method, the following equation is implemented:

$$X_{t_{i+1}} = X_{t_i} + \mu(t_i, X_{t_i})\Delta t + \sigma(t_i, X_{t_i})\Delta B_i \tag{8}$$

for all $i = 0, 1, \dots, k - 1$. In order to carry out the method computationally, it is necessary to know how to calculate ΔB_i . Since the partition is made up of equal intervals, the differences $\Delta B_i, i = 0, 1, \dots, k - 1$ have the same distribution, $\Delta B_i \sim N(0, \Delta t)$. Let η be a random variable with a standard normal distribution $\eta \sim N(0, 1)$. Then $\sqrt{\Delta t}\eta$ has a normal distribution with zero mean and variance Δt ; that is, $\sqrt{\Delta t}\eta \sim N(0, \Delta t)$. We use the following algorithm to implement the Euler-Maruyama method to find the approximate solutions of the stochastic differential equations:

Algorithm 1: Euler-Maruyama method

Define: $\mu(t, X_t), \sigma(t, X_t)$;
Input: $t_0, X_{t_0}, \Delta t, k$;
for $i := 0$ **to** $k - 1$ **do**
 $\eta_i \sim N(0, 1)$
 $X_{t_{i+1}} = X_{t_i} + \mu(t_i, X_{t_i})\Delta t + \sigma(t_i, X_{t_i})\sqrt{\Delta t}\eta_i$
 $t_{i+1} = t_i + \Delta t$
 print t_{i+1} and $X_{t_{i+1}}$

To implement in an analogous way the algorithm of the Euler-Maruyama method, previously described, for our proposed model, the respective discretization of the system of stochastic differential equation (1) must be carried out, which is given by:

$$\begin{aligned}
 S_{t_{i+1}} &= S_{t_i} - \left[\phi S_{t_i} - \psi C_{t_i} + \beta S_{t_i} \left(\frac{I_{t_i}^A + I_{t_i}^S}{N} \right) \right] \Delta t - \frac{S_{t_i} \sigma_1 I_{t_i}^A}{N} \sqrt{\Delta t} \eta_{1i} - \frac{S_{t_i} \sigma_2 I_{t_i}^S}{N} \sqrt{\Delta t} \eta_{2i} \\
 C_{t_{i+1}} &= C_{t_i} + [\phi S_{t_i} - \psi C_{t_i}] \Delta t \\
 E_{t_{i+1}} &= E_{t_i} + \left[\beta S_{t_i} \left(\frac{I_{t_i}^A + I_{t_i}^S}{N} \right) - \kappa E_{t_i} \right] \Delta t \\
 I_{t_{i+1}}^A &= I_{t_i}^A + \left[(1 - p) \kappa E_{t_i} - \omega I_{t_i}^A \right] \Delta t + \frac{S_{t_i} \sigma_1 I_{t_i}^A}{N} \sqrt{\Delta t} \eta_{1i} \\
 I_{t_{i+1}}^S &= I_{t_i}^S + \left[p \kappa E_{t_i} + (1 - \nu) \omega I_{t_i}^A - \xi I_{t_i}^S \right] \Delta t + \frac{S_{t_i} \sigma_2 I_{t_i}^S}{N} \sqrt{\Delta t} \eta_{2i} \\
 Q_{t_{i+1}} &= Q_{t_i} + [\xi I_{t_i}^S - \gamma Q_{t_i}] \Delta t \\
 H_{t_{i+1}} &= H_{t_i} + [q \gamma Q_{t_i} - ((1 - x) \alpha_R + x \alpha_D) H_{t_i}] \Delta t \\
 R_{t_{i+1}} &= R_{t_i} + [\nu \omega I_{t_i}^A + (1 - q) \gamma Q_{t_i} + (1 - x) \alpha_R H_{t_i}] \Delta t \\
 D_{t_{i+1}} &= D_{t_i} + [x \alpha_D H_{t_i}] \Delta t.
 \end{aligned} \tag{9}$$

3. Computational results from the statistical data: A case study from the city of Bogotá

In this section, we discuss the stochastic dynamics with environmental noises for the COVID-19 epidemic in Bogotá city. The data used in this article for the simulations, parameter adjustment, and analysis was selected from the publicly available database of the National Institute of Health of Colombia (or in Spanish: Instituto Nacional de Salud, INS) ([Instituto Nacional de Salud de Colombia, 2021](#)). This data source provides daily case numbers for both symptomatic and asymptomatic infected individuals without making a clear distinction between the two conditions. It also gives the report of recovered and deceased individuals. However, given that the selected epidemiological data often had irregularities, data cleaning was performed. Subsequently, the new database was filtered with respect to the dates of the analysis. In addition, it is necessary to have the total number of inhabitants in the city of Bogotá D.C. for the mathematical modelling. The population projection of the National Administrative Department of Statistics (or in Spanish: Departamento Administrativo Nacional de Estadística, DANE) ([Administrativo Nacional de Estadística, 2018](#)), estimates that for 2020 the total population in the respective city is 7743 955 persons. Finally, the entire methodological procedure was developed in Python 3.7.

For the proposed model (1) model, we have taken into account the first 200 days of the pandemic in the city of Bogota. The first contagion report occurred on March 6, 2020 ([Ministerio de Salud y Protección Social, 2020](#)) (*day 1*) and the end date of the analysis was set for September 22, 2020 (*day 200*). The calibration of the model was carried out with the infected, recovered, and deceased data of the first 100 days, that is, from March 6, 2020, to June 14, 2020 (*day 100*). It is planned for the period from June 15, 2020 to September 22, 2020. The lockdown began on March 24, 2020 ([Presidencia de la República de Colombia, 2020a, 2020b](#)) (*day 18*) and the partial opening began on May 11, 2020 ([Presidencia de la República de Colombia, 2020a, 2020b](#)) (*day 66*). By September 22, 2020, the following was reported in Colombia: 777 537 infected, 650 801 recovered, 24 570 deaths, and 3460 714 processed samples. And for the city of Bogotá D.C, the following was reported: 256,162 infected, 213,956 recovered, and 6,610 deaths (see [Table 2](#)). Regarding the daily report of infected in the same city, its maximum value or peak occurred around August 6, 2020, with 6068 infected cases (see [Figs. 2 and 3](#)).

3.1. Calibration of the model

The calibration of the proposed model is carried out in order to find a set of parameters so that the model has a good description of the behavior of the system dynamics, i.e. so that the model fits the data as well as possible and allows us to make the desired projections. The search for the parameters can be carried out by means of different optimization algorithms. The non-linear least squares method of trust region was the one chosen (see ([Sáez and Rittmann, 1992](#))), which will be explained later on. Subsequently, the predictions of the model are compared with the real data available for the city of Bogotá.

It is necessary to find the values of the parameters that fit the model according to the available data and satisfy a certain set of constraints to ensure global convergence and consistency of the parameters. We present the nonlinear least squares optimization method of trust region described in ([Kloeden and Platen, 1992](#)) for the adjustment of the parameters of the proposed model. Let $\theta = (\theta_1, \dots, \theta_m)$ be a multidimensional parameter with a neighborhood defined by $B_{\Delta k}(p) = \{\theta_k \in \mathbb{R} : \|\theta - p\| \leq \Delta k\}$ for $1 \leq k \leq m$, where Δk is the radius of the trusted region. Consider $f : \mathbb{R}^n \rightarrow \mathbb{R}$ a function that depends on the parameter θ .

Given some points $(x_1, y_1) \dots (x_n, y_n)$ the residual sum of squares is defined as:

Table 2
Statistical description of the data for the city of Bogotá D.C

	Infected	Accumulated infected	Recovered	Accumulated recovered	Dead	Accumulated dead
Mean	1300.213 198	65 121.263 959	1138.063 830	42 899.111 702	36.927 374	1972.854 749
Std	1419.070 253	83 781.640 668	1593.048 383	62 834.182 549	35.517 685	2266.236 273
Min	1.000 000	1.000 000	1.000 000	1.000 000	1.000 000	1.000 000
25%	112.000 000	2390.000 000	61.250 000	1309.750 000	5.000 000	172.500 000
50%	552.000 000	16 853.000 000	372.500 000	9367.000 000	25.000 000	639.000 000
75%	2174.000 000	113 841.000 000	1734.500 000	58 568.250 000	65.500 000	3965.000 000
Max	6068.000 000	256 142.000 000	7354.000 000	213 956.000 000	122.000 000	6610.000 000

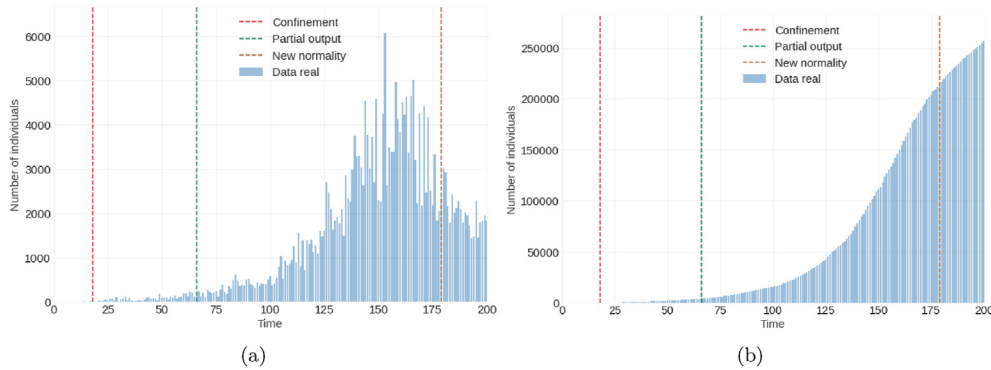


Fig. 2. Graph (a) on left denotes number of infected individuals(incidence); Graph (b) on right denotes cumulative number of infected individuals.

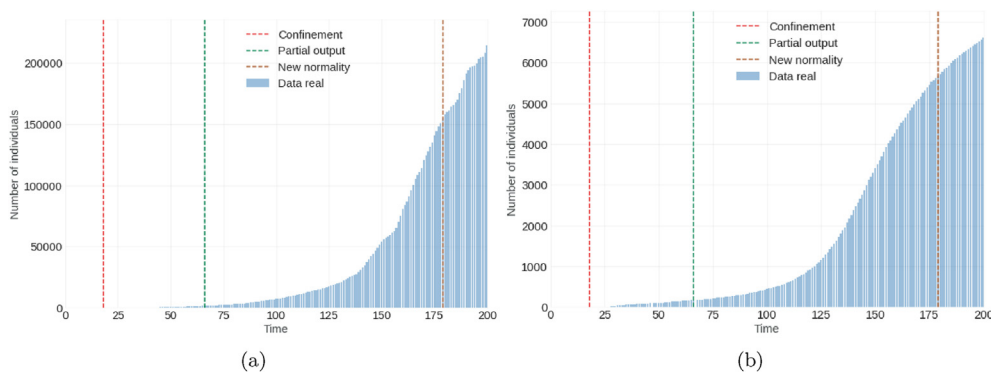


Fig. 3. Graph (a) on the left denotes cumulative number of individuals recovered; Graph (b) on the right denotes cumulative number of deceased individuals.

$$g(\theta) = \|y - f(x; \theta)\|_2^2 = \|R(\theta)\|_2^2 \tag{10}$$

where $g(\theta)$ is the objective function. The information from the objective function in the trust region methods is used to form a model $m_k(7)$, which, near the current point θ_k , have as much as possible a behavior similar to that of the objective function.

$$m_k(p) = \|R_k(\theta_k) + J_k p\|_2^2 \tag{11}$$

where the Hessian is $\nabla^2 g \approx J^T J$. Since the model may not be a good approximation of the objective function, when taking points θ away from θ_k , the search for the minimum m_k that must be restricted to points at $B_{\Delta k}$. So, in each iteration of this methodology, the following sub-problem is solved;

$$\min_p m_k(p),$$

subject to $\|p\| \leq \Delta k$.

Table 3
Adjusted drift parameters.

Parameter	Value	Units	Standard error	Bounds	Reference
β	1.063 147	day	2.965 016	[0, 1.4]	de-Camino-Beck (2020); Kumari et al. (2020)
ω	0.476 190	days ⁻¹	13.234 271	[1/10, 1/2.1]	Mejía Becerra (2020); Imperial College London (2020)
ξ	0.200 000	days ⁻¹	1.614 248	[1/2, 1]	Harvard (2020); Belkhiria and Nascimento (2020)
α_R	0.105 263	days ⁻¹	6.379 961	[1/12, 1/9.5]	Jafari. et al. (2021)
α_D	0.131 579	days ⁻¹	4.1 526E+2	[1/14, 1/7.6]	Jafari. et al. (2021)
ν	0.731 120	—	12.446 559	(0, 1)	Imperial College London (2020)
χ	0.900 000	—	3.1 375E+3	(0, 1)	Imperial College London (2020)

Given the data on infected, recovered, and deceased, and the information on the parameters obtained from a meta-analysis of the epidemiological and health entities in charge of Bogotá, the parameters of the model were fitted with the non-linear least squares procedure using the trust region algorithm. In Table 3 and Table 4, we can observe the fixed values and those obtained from the model adjustment respectively, and also in the references column we can find the literature, which helped us to establish the fixed and initial values to carry out the optimization process. Note that the standard error for the parameters α_D and χ are too high, which possibly indicate that there is too much variability of the number of deaths after COVID-19 hospitalized them: there could be days with few casualties and others with too many deaths after being hospitalized by COVID-19.

3.2. Model projections

Fig. 4 and Fig. 5 generally illustrate the contrast between the real dynamics and the projections of the proposed model in the first mitigation measures for the pandemic for the SARS-CoV-2 coronavirus. Fig. 4a shows the analysis of the daily report of individuals infected by the virus, that is, the incidence of infected. In contrast, Fig. 4b shows the analysis of the number of accumulated infected. Fig. 5a and b represents the analysis of accumulated recovered and accumulated deceased individuals, respectively.

In the following figures, the actual data is represented by the yellow line. The gray lines are the 20 simulations represents results from the stochastic model subject to parameter adjustment by the trust region method, and the blue line is the drift of the model. Likewise, the graphs specify the analysis period, that is, the period in which the parameters were projected and adjusted. They also indicate the periods of confinement and partial opening.

Table 4
Fixed parameters.

Parameter	Value	Units	Reference
κ	1/4.6	days ⁻¹	Mejía Becerra (2020); Imperial College London (2020)
γ	1/8	days ⁻¹	(Harvard, 2020), (Belkhiria & Nascimento., 2020)
p	0.100	—	Mejía Becerra (2020); He et al. (2021)
q	0.190	—	Ohajunwa et al. (2020); Belkhiria and Nascimento (2020)
a	0.934	—	Assumption
b	0.011	—	Assumption
ϵ	5.000	—	Assumption
t_a	18	days	Presidencia de la República de Colombia (2020a, 2020b)
t_b	66	days	Presidencia de la República de Colombia (2020a, 2020b)

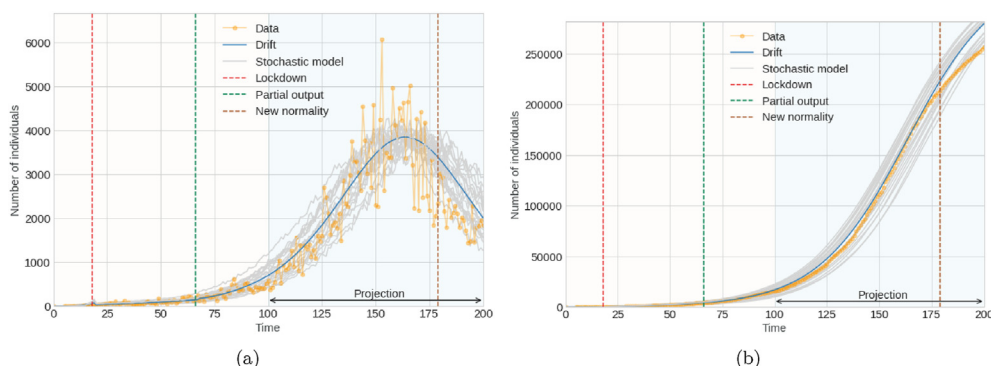


Fig. 4. Graph (a) on left denotes number of infected individuals (incidence); Graph (b) on right denotes cumulative number of infected individuals.

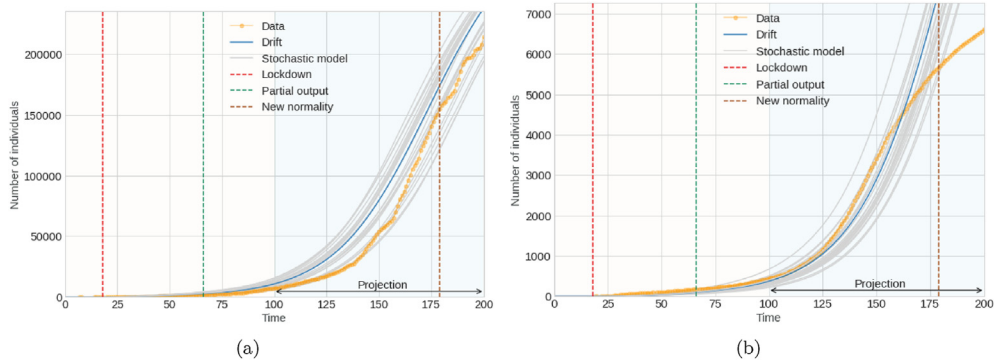


Fig. 5. Graph (a) on left denotes cumulative number of individuals recovered; Graph (b) on right denotes cumulative number of deceased individuals.

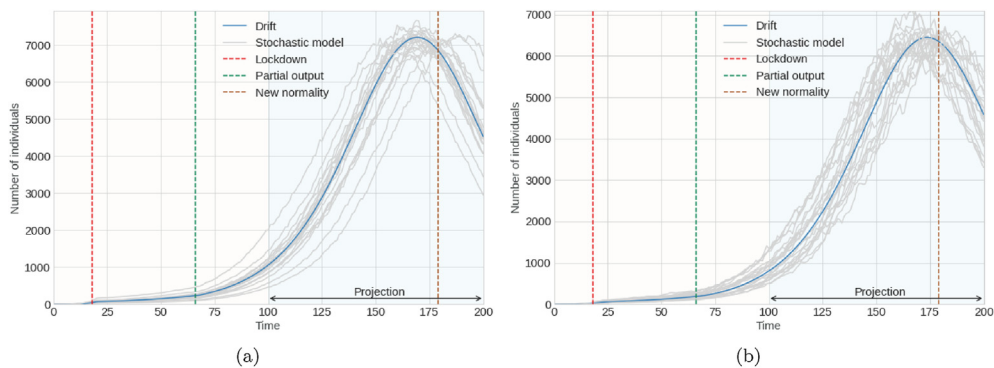


Fig. 6. Graph (a) on left denotes number of asymptomatic infected individuals (prevalence); Graph (b) on right denotes number of symptomatic infected individuals (prevalence).

The mean square logarithmic error (MSLE) is calculated as an evaluation measure of the proposed stochastic model. The MSLE only considers the relative difference between the actual and the projected value. This measure treats minor differences between small actual and forecasted values, roughly the same way as significant differences between large actual and forecasted values. The cause of errors is significantly penalized than small ones in those cases where the target value range is large. So, since the drift is the stochastic mean, when evaluated against the actual data with the MSLE measure, this gives a cursory idea of how reliable the model is. For the projections of the incidence of infected, an MSLE of 0.090 is obtained, and for the accumulated infected, an MSLE of 0.005, while for the predictions of accumulated recovered, an MSLE of 0.157 and for accumulated deaths an MSLE of 0.062. Therefore, both the graphs and the superficial or general measurement of the MSLE model infer that the projections of the recovered are not very favorable compared to the other cases.

Fig. 6 and Fig. 7 are not contrasting graphs like the previous ones. One reason could be the difficulty of obtaining the prevalence data for the cases mentioned below. However, as can be seen, the graphs represent the 20 simulations of the stochastic model, the drift of the model given the adjustment of the parameters and with intensity constants $\sigma_1 = 0.2$ and $\sigma_2 = 0.4$ and the dates of the mitigation strategies chosen by the national government. Fig. 6a and b show possible projections of the stochastic model for asymptomatic and symptomatic infected individuals, respectively. In contrast, Fig. 7a represent simulations of the prevalence of hospitalized individuals, and Fig. 7b represents these changes in the stochastic dynamics of susceptible individuals.

In our model, we obtain the basic reproduction number for the proposed model (theoretically calculated in (Ohajunwa et al., 2020)) given by

$$R_0^{Cov} = \beta \left(\frac{2p}{\xi} + \frac{1-p}{\omega} + \frac{(1-v)(1-p)}{\xi} \right)$$

Also, from the parameters given in Table 4, we note that the basic reproduction number for COVID-19 in Bogotá has normal distribution with median $R_0^{Cov} = 4.34$ and variance 1.79 (standard deviation 1.34) given by:

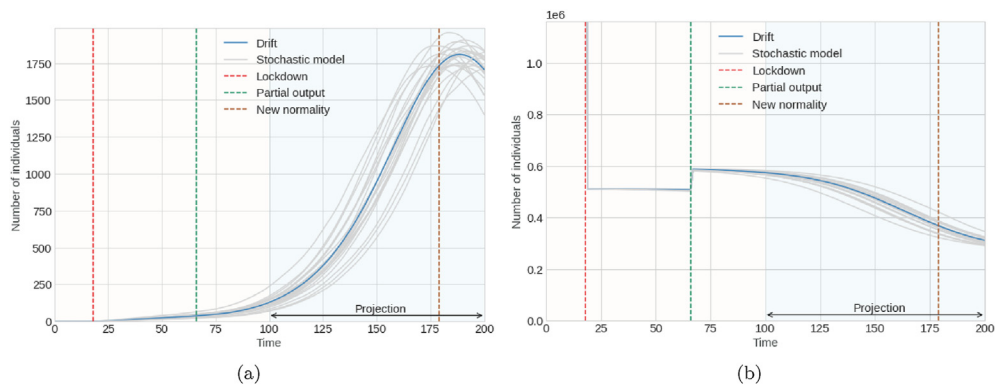


Fig. 7. Graph (a) on left denotes number of people hospitalized (prevalence); Graph (b) on right denotes number of susceptible people.

$$\frac{\sigma_1^2((1-v)^2(1-p)+p)^2}{2\xi^3} + \frac{\sigma_2^2 p^2}{2\xi^3} + \frac{\sigma_2^2(1-p)^2}{2\omega^3} + \frac{2\sigma_2^2 p(1-p)}{(\xi+\omega)\xi\omega}$$

suggesting that COVID-19 could be an *endemic* for Bogotá. This may also be due to the μ variant B.1.621 identified in Colombia and recognized by WHO in the list of variants that reported 25% of new incidence cases in Colombia (see (Khan Burki, 2021; Shultz et al., 2021)).

4. Conclusion

In this work, a new stochastic epidemiological model is proposed that accounts for randomness in the dynamics of COVID-19. The model provides a perspective of the possible scope for the spread of COVID-19 with random behaviors in the population. We were able to show the importance of including stochasticity within the population that models the real data from Colombian city of Bogotá. Our computational simulations suggest that the proposed model is reliable and robust in projecting the number of infected, recovered and deceased individuals during interventions such as confinement and partial opening. The results from this work can be used to inform data-driven decision making for improving public health in Colombia.

While the model seems to capture the data from the Colombian city of Bogotá well, it is potentially limited by the lack of prevalence data corresponding to registration of individuals who present a certain characteristic of the disease over a period of time, including susceptible, exposed, hospitalized and quarantined sub-populations, as well as the complete registry of the data of asymptomatic infected. The proposed model explains the containment of the pandemic by strict lock-down and decline in number of incidence cases and death. This work also opens a way to explore newer compartment models one can take into account including other factors such as migrations, demographics, spatial and vaccination effects. One can also compare the proposed model against reliable data from another country. Another potential area of research includes using these models in conjunction with machine learning and deep learning to obtain optimal parameters. Another area that can be considered is to study to what extent the results can be generalized to the overall population and across different geographical areas as well as patterns of contact between and within groups and in different social settings. All these will be considered in forthcoming papers.

Availability of data and material

Not applicable. All data generated or analyzed during this study are included in this manuscript.

Funding

This work was financial support by Directorate-Bogotá campus (DIB), Universidad Nacional de Colombia under the project No. 50803.

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.

Acknowledgements

The excellent comments of the anonymous reviewers are greatly acknowledged and have helped a lot in improving the quality of the paper.

References

- de-Camino-Beck, T. (2020). A modified SEIR model with confinement and lockdown of COVID-19 for Costa Rica. In *medRxiv*.
- Administrativo Nacional de Estadística, D. (2018). Proyecciones de población a nivel departamental. Periodo 2018 - 2050. url: <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion>.
- Alberto, M., Gong, Z., & Giordano, S. (2021). Modelling, prediction and design of COVID-19 lockdowns by stringency and duration. In *Scientific reports*, 11.1 pp. 1–13).
- Amouch, M., & Karim, N. (2021). Modeling the dynamic of COVID-19 with different types of transmissions. In *Chaos, solitons & fractals* (p. 111188).
- Belkhiria, & Nascimento, J L. W do, Jr. (2020). COVID-19 pandemics modeling with modified deterministic SEIR, social distancing, and age stratification. The effect of vertical confinement and release in Brazil. In *Plos one* <https://doi.org/10.1371/journal.pone.0237627>. url:.
- Benítez, M. A., et al. (2020). Responses to COVID-19 in five Latin American countries. In *Health policy and technology*, 9.4 pp. 525–559). <https://doi.org/10.1016/j.hlpt.2020.08.014>
- Brauer, F., & Castillo-Chavez, C. (2012). *Mathematical models in population biology and epidemiology*. Springer.
- Cai, S., Cai, Y., & Mao, X. (2019). A stochastic differential equation SIS epidemic model with two independent Brownian motions. In *Journal of mathematical analysis and applications*, 474.2 pp. 1536–1550).
- Cárdenas, M., & Martínez, H. (2020). COVID-19 in Colombia. *Impact and Policy Responses*. In Center for Global Development.
- Chung, K. L., & Williams, R. J. (2014). *Introduction to stochastic integration*. Springer.
- De la Hoz-Restrepo, F., et al. (2020). Is Colombia an example of successful containment of the 2020 COVID-19 pandemic? A critical analysis of the epidemiological data, March to July 2020. In *International journal of infectious diseases* (Vol. 99, pp. 522–529).
- Gray, A., et al. (2011). A stochastic differential equation SIS epidemic model. In *SIAM journal on applied mathematics*, 71.3 pp. 876–902).
- Harvard, H. U. (2020). Health Health Information and Medical Information, if you've been exposed to the coronavirus [Online; accessed 5-May-2021] <https://www.health.harvard.edu/diseases-and-conditions/if-youve-been-exposed-to-the-coronavirus#:~:text=The%5C%20most%5C%20recent%5C%20CDC%5C%20guidance,days%5C%20since%5C%20your%5C%20symptoms%5C%20began>.
- He, J., et al. (2021). In *Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis* (Vol. 93, pp. 820–830). Wiley J Med Virol. <https://doi.org/10.1002/jmv.26326>.
- Imperial College London. (2020). MRC centre for global infectious disease analysis. url: <https://mrc-ide.github.io/global-Imic-reports/parameters.html>.
- Instituto Nacional de Salud de Colombia. Coronavirus (COVID - 19) en Colombia. url: <https://www.ins.gov.co/Noticias/Paginas/Coronavirus.aspx>.
- Instituto Nacional de Salud de Colombia. (2021). Boletines casos COVID19 Colombia. url: <https://www.ins.gov.co/BoletinesCasosCOVID19Colombia/Forms/AllItems.aspx>.
- Jafari, Y., Rees, E. M., & Nightingale, E. S. (2021). COVID-19 length of hospital stay: A systematic review and data synthesis. In *BMC med* (Vol. 18, p. 270). <https://doi.org/10.1186/s12916-020-01726-3>
- Khan Burki, T. (2021). The race between vaccination and evolution of COVID-19 variants. In *The lancet respiratory medicine*.
- Klebaner, F. C. (2012). *Introduction to stochastic calculus with applications*. World Scientific Publishing Company.
- Kloeden, P. E., & Platen, E. (1992). *Numerical solution of stochastic differential equations*. Springer-Verlag Berlin Heidelberg.
- Komarova, N. L., Azizi, A., & Wodarz, D. (2021). Network models and the interpretation of prolonged infection plateaus in the COVID19 pandemic. In *Epidemics* (Vol. 35, p. 100463).
- Kumari, P., Singh, H. P., & Singh, S. (2020). SEIAQRDT model for the spread of novel coronavirus (COVID-19): A case study in India. In *Applied Intelligence*. <https://doi.org/10.1007/s10489-020-01929-4>
- Leung, K. Y., et al. (2018). Individual preventive social distancing during an epidemic may have negative population-level outcomes. In *Journal of the royal society interface*, 15.145 p. 20180296).
- Lin, Q., et al. (2020). A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. In *International journal of infectious diseases* (Vol. 93, pp. 211–216).
- Maier, B. F., & Brockmann, D. (2020). Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. In *Science*, 368. 6492 pp. 742–746).
- Maruyama, G. (1955). Continuous Markov processes and stochastic equations. In *Rendiconti del Circolo Matematico di Palermo*, 4.1 p. 48).
- Mejía Becerra, Juan Diego (2020). In *Modelación Matemática de la Propagación del SARS-CoV-2 en la Ciudad de Bogotá Tercera Versión*. Secretaría Distrital de Salud de Bogotá.
- Ministerio de Salud y Protección Social. (2020). Colombia confirma su primer caso de COVID-19. url: <https://www.minsalud.gov.co/Paginas/Colombia-confirma-su-primer-caso-de-COVID-19.aspx>.
- Musa, S. S., et al. (2021). Mathematical modeling of COVID-19 epidemic with effect of awareness programs. In *Infectious disease modelling* (Vol. 6, pp. 448–460).
- Ohajunwa, C., Kumar, K., & Seshaiyer, P. (2020). Mathematical modeling, analysis, and simulation of the COVID-19 pandemic with explicit and implicit behavioral changes. In *Computational and Mathematical Biophysics*, 8 pp. 216–232). De Gruyter Open.
- Oraby, T., et al. (2021). Modeling the effect of lockdown timing as a COVID-19 control measure in countries with differing social contacts. In *Scientific reports*, 11.1 pp. 1–13).
- Prem, K., et al. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. In *The lancet public health* (Vol. 5, p. 5). [https://doi.org/10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6)
- Presidencia de la República de Colombia. (2020a). Estas son las actividades productivas que empezarán a reactivarse a partir del 11 de mayo, con gradualidad y protocolos. url: <https://id.presidencia.gov.co/Paginas/prensa/2020/Presidente-Duque-anuncia-Aislamiento-Preventivo-Obligatorio-todo-pais-a-partir-proximo-martes-24-marzo-a-la-23-59-ho-200320.aspx>.
- Presidencia de la República de Colombia. (2020b). Presidente Duque anuncia Aislamiento Preventivo Obligatorio, en todo el país, a partir del próximo martes 24 de marzo, a las 23:59 horas, hasta el 13 de abril, a las cero horas. url: <https://id.presidencia.gov.co/Paginas/prensa/2020/Presidente-Duque-anuncia-Aislamiento-Preventivo-Obligatorio-todo-pais-a-partir-proximo-martes-24-marzo-a-la-23-59-ho-200320.aspx>.
- Rădulescu, A., Williams, C., & Cavanagh, K. (2020). Management strategies in a SEIR-type model of COVID 19 community spread. In *Science* (Vol. 10, p. 1). <https://doi.org/10.1038/s41598-020-77628-4>
- Ríos-Gutiérrez, A., Torres, S., & Arunachalam, V. (2021). Studies on the basic reproduction number in stochastic epidemic models with random perturbations. In *Advances in difference equations 2021* (Vol. 1, pp. 1–24). <https://doi.org/10.1186/s13662-021-03445-2>
- Sáez, P. B., & Rittmann, B. E. (1992). Model-parameter estimation using least squares. In *Water research*, 26.6 pp. 789–796).
- Shultz, J. M., et al. (2021). Complex correlates of Colombia's COVID-19 surge. In *The lancet regional health—americas* (Vol. 3).
- Sohrabi, C., et al. (2020). World health organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). In *International journal of surgery* (Vol. 76, pp. 71–76). [https://doi.org/10.1016/S0140-6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9)

- Wang, C., et al. (2020). A novel coronavirus outbreak of global health concerns. In *Lancet*, 395.10223 pp. 470–473). [https://doi.org/10.1016/S0140-6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9)
- World Health Organization. (March 2020). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020>. 2020.
- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. In *The lancet*, 395.10225 pp. 689–697).