

## RESEARCH ARTICLE

# An updated estimation approach for SEIR models with stochastic perturbations: Application to COVID-19 data in Bogotá

Andrés Ríos-Gutiérrez<sup>1,2</sup> , Soledad Torres<sup>3</sup> , Viswanathan Arunachalam<sup>1</sup> \*

**1** Department of Statistics, Universidad Nacional de Colombia, Bogotá, Colombia, **2** School of Mathematics, Universidad Industrial de Santander, Bucaramanga, Colombia, **3** CIMFAV - Facultad de Ingeniería, Universidad de Valparaíso, Valparaíso, Chile

 These authors contributed equally to this work.

\* [varunachalam@unal.edu.co](mailto:varunachalam@unal.edu.co)



## Abstract

This paper studies the updated estimation method for estimating the transmission rate changes over time. The models for the population dynamics under SEIR epidemic models with stochastic perturbations are analysed the dynamics of the COVID-19 pandemic in Bogotá, Colombia. We performed computational experiments to interpret COVID-19 dynamics using actual data for the proposed models. We estimate the model parameters and updated their estimates for reported infected and recovered data.

## OPEN ACCESS

**Citation:** Ríos-Gutiérrez A, Torres S, Arunachalam V (2023) An updated estimation approach for SEIR models with stochastic perturbations: Application to COVID-19 data in Bogotá. PLoS ONE 18(8): e0285624. <https://doi.org/10.1371/journal.pone.0285624>

**Editor:** Hana Maria Dobrovolny, Texas Christian University, UNITED STATES

**Received:** September 22, 2022

**Accepted:** April 26, 2023

**Published:** August 21, 2023

**Copyright:** © 2023 Ríos-Gutiérrez et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The necessary code and data for achieving these results are available online at: <https://github.com/andressriosg/withsoledad>.

**Funding:** The author VA acknowledge support from the DIB, Universidad Nacional de Colombia project No.57809 and MATH-AmSud 18 MATH-07 SaSMoTiDep Project. The author ST acknowledge support from Basal Project FB210005, and Fondecyt 1221373 and 1230807.

## Introduction

The Latin American region is one of the regions that has reported more COVID-19 infection cases. At the end of the year, 2019, the World Health Organization (WHO) said the outbreak started in Wuhan, China. This outbreak quickly spread to more than 50 countries in one month. On January 30<sup>th</sup>/2020, the WHO declared the COVID-19 epidemic a Public Health Emergency of International Concern (PHEIC) [1]. European countries such as Spain, France, and Italy have had a significant number of deaths and a high number of infected cases. In March 6<sup>th</sup>/2020, the first case of COVID-19 coronavirus was confirmed in Colombia [2]. Colombia has implemented many emergency measures in response to the coronavirus outbreak, including strict lock-downs, PCR testing capacity, contact tracing, and augmenting ICU capacity in the hospitals. In particular, Colombia is one of the top ten countries globally regarding the registered number of infections for more than 2 million cases and more than 58,974 deaths since March 2020, regarding the data of the COVID-19 of the Instituto Nacional de Salud Colombia (INS) [3] (See Fig 1). In Colombia, national and local governments have taken measures to control the spread of infections, such as lockdowns, restrictions movements, including closing airports, etc. [4, 5]. However, there was still a rapid spread of the virus in Colombia, even when the vaccination in Colombia started on February 17<sup>th</sup> 2021 and, in May 22<sup>nd</sup> 2021 there were 9,325,861 vaccinated people. [6]

Given the conditions of contagion transmission and spread in the country, it is necessary to establish when the rise of cases is expected to be high and when the isolation and modality

**Competing interests:** The authors declare that they have no competing interests.

restriction measures can be repealed. To mention a few models, SIR (susceptible, infected, and recovered), SIS (susceptible and infected), and SEIR (susceptible, exposed, infected, and recovered). The first epidemiological mathematical model in 1927 by Kermack and McKendrick in [7], the SIR model. In this model, three types of individuals are distinguished: susceptible in the time  $t$  ( $S(t)$ ), infected in the time  $t$  ( $I(t)$ ), and removable in the time  $t$  ( $R(t)$ ). Susceptible individuals are prone to contracting a disease by having infectious contact with these infected who may present symptoms. Removable individuals cease to be infected, or either die or have immunity against the disease.

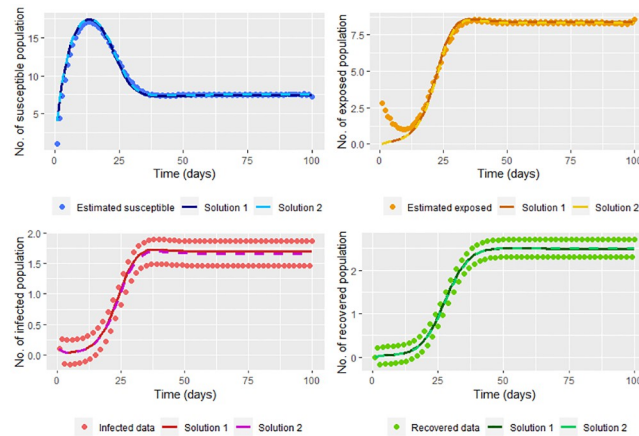
In the SIR model,  $\beta$  is the *transmission rate*. Thus,  $\beta I(t)S(t)$  is the total of  $S(t)$  susceptible who acquire the infectious agent when having contact with  $I(t)$  infectious population on time  $t$ .  $1/\gamma$  is the *recovered time* ( $\gamma$  is the *recovered rate*), that is, the time that takes an infected individual to recover in a removable individual. This way,  $\gamma I(t)$  is removable of  $I(t)$  infected individuals on time  $t$ . Something intrinsic to the SIR model does not consider the *exposed population*, who are the individuals that have the infectious agent, but they can not spread the epidemic. The notation is  $E(t)$  as the number of exposed populations in the day  $t$ . That population can be the infection agent (virus, bacteria, etc.) but is not infectious until they have completed the entire incubation. This time is time that takes an individual to recover from a contagious individual, which we denote by  $1/v$ . In COVID-19, there is also an incubation time from 2 to 14 days [8]. For this reason, we consider a more general model: the SEIR model [9]. In the SEIR model,  $\beta I(t)S(t)$  and  $\gamma I(t)$  are interpreted similarly to the SIR model. In this model,  $v$  is the incubation rate. The SIR and SEIR models assume the total population is constant. For example, the SEIR model is

$$N(t) = S(t) + E(t) + I(t) + R(t)$$

The modified SEIR model is the *SEIR model with demographics*, where  $\Lambda$  represents the *influence rate*, that is, the average number of new susceptible populations per unit of time [10]. The *emigration rate* is denoted as  $\mu$ , and  $\gamma$  is the *recovered rate*. Therefore,  $\gamma I(t)$  is the total recovered of  $I(t)$  infected individuals on time  $t$ . The equation of the SEIR model with demographics is the ordinary differential system 1.

$$\begin{cases} \frac{dS(t)}{dt} = \Lambda - \beta I(t)S(t) - \mu S(t) \\ \frac{dE(t)}{dt} = \beta I(t)S(t) - vE(t) - \mu E(t) \\ \frac{dI(t)}{dt} = vE(t) - \gamma I(t) - \mu I(t) \\ \frac{dR(t)}{dt} = \gamma I(t) - \mu R(t) \end{cases} \tag{1}$$

Recently, research articles have been studied using different approaches to model COVID-19, to mention a few [11–15]. They focus on minimizing the sum of squares based only on the infected population data in taking the parameters of other papers. Statistically, it should be used all the types of populations under an epidemic and a methodology to estimate the parameters of the model. Actually [16] is highlighted that “it is difficult to consider all possible interactions between interventions in the same model and find parameters close to reality through simulations”. In this way, we must search for a model which includes all the populations under the epidemic and estimates the parameters based on as much information as possible. In particular, to Bogotá (Colombia), there are previously published papers as [17–19] whose focus is



**Fig 1. COVID-19 infected data from Bogota D.C. during the first 385 days.**

<https://doi.org/10.1371/journal.pone.0285624.g001>

taking the parameters under which a good fit is observed. These papers are based on SEIR models with types of population, for instance, asymptomatic or hospitalized individuals, but they do not estimate the susceptible and exposed population. However, the fit is only on the number of reported cases of COVID-19, leaving aside the recovered population, whose data is being reported by the [3]. The main objective of this paper is to provide three different methods to estimate parameters on models based on epidemics, taking as an example the SEIR model: (i) minimizing the loss function considering infected and recovered data; (ii) using a data update approach; and (iii) using a stochastic infection rate. We also give a possible implementation to actual data, in this case, considering the COVID-19 data from Bogotá—one of the methods to estimate the susceptible and exposed population: using the data update estimation.

This paper describes methods for estimating SEIR-type models for COVID-19 data in Bogota. The paper is organized as follows: The section “A state of the art on the parameter estimation on epidemic models” proposes two models based on the minimization of loss functions based on infected population  $I(t)$  and recovered population  $R(t)$  data. The section on “Estimation of the SEIR MODEL from real data” is devoted to parameter estimation based on the susceptible  $S$  and exposed  $E$  population, which has been previously estimated using real data. Additionally, we update the parameters  $\beta$ ,  $\nu$ , and  $\gamma$  as a function of time. The section on “Estimation of the SEIR MODEL from real data” presents an SEIR model with random perturbations, i.e., where we assume that the parameter  $\beta$  is random. We also include the corresponding parameter estimation method for real data for this model. Finally, the last section concludes the paper with future perspectives and the advantages and disadvantages of each of the proposed methodologies.

## A state of the art on the parameter estimation on epidemic models

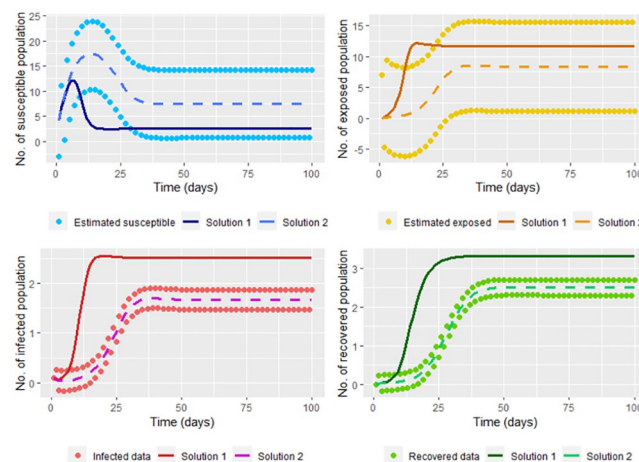
### State-of-art

This paper establishes an estimation methodology that could improve the forecasts for the populations in an epidemic under the SEIR model. For example, we take COVID-19 data from Bogota during the first 385 days of the pandemic. According to the test, we assume that the number of infected populations is the number of individuals diagnosed as infected. The number recovered population is the number of individuals who, after being positive for COVID-

19, had a negative test reported each day or who have been 14 days without symptoms. We do not consider the recovered data. The cumulative number of tests reported as negative measures individuals could be reported as newly infected in the following days when the patient retests once they present new symptoms. Data is taken from the public database of the INS [3]. In addition, we consider that the infection rate is not the same for all 385 days since most people from Bogota were initially isolated by control and prevention measures, which can be searched in [5]. After around 150 days, there were economic reactivation and Christmas holidays, so the infection rate differed (see Fig 2). We have **six periods** of time, each with different infection rates, which were chosen regarding to the rules and laws established by the Colombian Government and mayor's office of Bogota. These points do not coincide necessarily with the inflection points of the infection function. We assume that the infection rate of each period changed according to the measures of mitigation and control against COVID-19 by the Bogota D.C. city [5]. Some of these points coincide with some peaks and valleys found in the graph of the smoothed infected data (see Fig 2). On the other hand, we do not consider variants of the disease, due to the available database does not have of the type of variant for each infected individual. The intervals are noted by  $\tau_1, \dots, \tau_6$ . In the Fig 2, we smoothed the data using the function `frfast` from `npregfast` package in R.

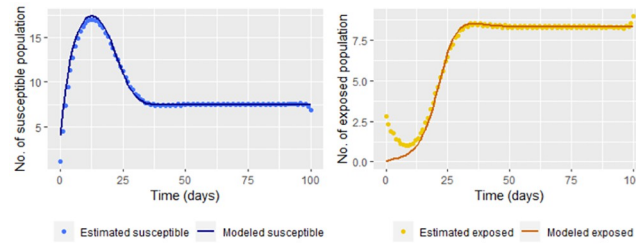
We propose to study the COVID-19 over-estimation methods using the system 1. There are different statistical methods of estimating parameters for the deterministic models, namely, a trial-and-error method, the use of computational algorithms for minimizing the sum of squares (see [20–25] or on the available data [26]). Also, it is worth mentioning that the papers [27–31] consider the parameters as a function of time  $t$  and estimate their values. However, some papers are based only on infected data, which drives possible flawed estimations on another type of population under an epidemic (Fig 3). Other papers, as [32, 33] suggest an overestimation of the infected population or even that the models with differential equations do not work to predict the infected population.

We use the maximum likelihood method to estimate the parameters using COVID-19 data from Bogota under the SEIR model with the regression model given by the Eq 2. We take  $\Lambda = (9.42/1000)$ ,  $\bar{p} = 73,660$  with  $\bar{p}$  as the average of the projected population of Bogota from march 6<sup>th</sup> of 2020 to march 21<sup>th</sup> of 2021 according to DANE, which is the Official Statistical Office from Colombia, the birth  $\Lambda = 9.42/1000$  and the death rate  $\mu = 4/1000$  (see [34]). We



**Fig 2.** COVID-19 data from Bogota D.C. during of the first 385 days.

<https://doi.org/10.1371/journal.pone.0285624.g002>



**Fig 3. Estimated infected and recovered population of the 4 model with  $\hat{\delta}_1^2 = \sum_{j=1}^n (I(t_j) - \phi_3(t_j))^2/n$  for each considered interval  $I_r, r = 1, \dots, 6$ , using the `mle2` function (parameters on the Table 1).**

<https://doi.org/10.1371/journal.pone.0285624.g003>

note  $A_k(t)$  with  $k = 1, \dots, 4$  as  $S(t), E(t), I(t)$  and  $R(t)$ , respectively, on 2. In addition  $\phi_k(t), k = 1, \dots, 4$  is noted on 2 as the regression function which corresponds to the approximate solution of the system 1 for  $S(t), E(t), I(t)$  and  $R(t)$ , respectively. We do not have the analytic solution of the system 1 on 2, therefore the approximate solution refers to the solution according to numerical methods according to `ode` function from `deSolve` package in R.

$$A_k(t) = \phi_k(t) + \epsilon_k, \epsilon_k \sim N(0, \delta^2). \tag{2}$$

On 2 we propose to estimate  $\delta^2$  as the variance given by

$$\hat{\delta}^2 = \sum_{k=1}^4 \sum_{j=1}^n \frac{(A_k(t_j) - \phi_k(t_j))^2}{n}, \tag{3}$$

thus, we use  $A_k(t_j) \sim N(\phi_k(t_j), \hat{\delta}^2), j = 1, \dots, n$  for the parameter estimation (see [35]). To this extent, we add noise for each of the variables  $A_k(t_j)$  with  $k = 1, \dots, 4$  to make a maximum likelihood estimation. We wish to mention that the recent papers [21, 23, 25] fit the models only to the infected data. Following that focus and the notation  $\phi_k(t_j), A_k(t_j)$ , we re-write the model 2 as the equation given by

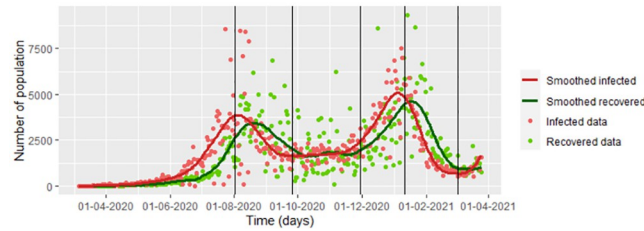
**Model 1:**  $A_k(t) = \phi_k(t) + \epsilon_k, \epsilon_k \sim N(0, \delta_1^2).$  (4)

with  $\hat{\delta}_1^2 = \sum_{j=1}^n (I(t_j) - \phi_3(t_j))^2/n$  (see Fig 3), which is a model with noise only for  $I(t)$ . We use `mle2` function from `bbmle` package from R, to establish the parameters for each period. We take six periods instead of 8 because intervals 2 and 3 and 4 and 5 do not have inflection points for the infected population. These parameters can be observed in the Table 1

**Table 1. Estimated parameters of the 4 model with  $\hat{\delta}_1^2 = \sum_{j=1}^n (I(t_j) - \phi_3(t_j))^2/n$ , for each interval  $I_r, r = 1, \dots, 6$ . We use the `mle2` function.**

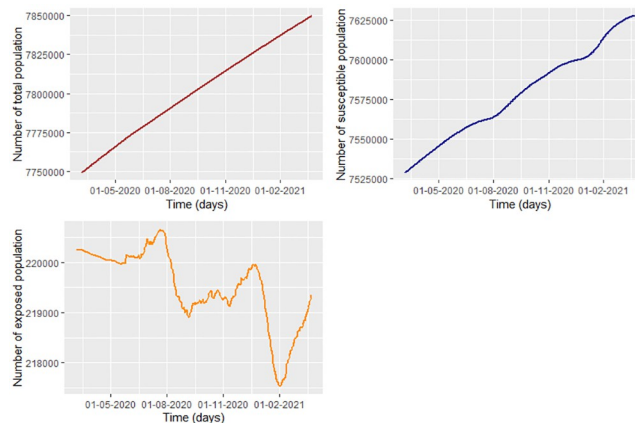
Period of time ( $\tau_s$ )	$\beta_s, \text{days}^{-1}$	$\nu_s, \text{days}^{-1}$	$\gamma_s, \text{days}^{-1}$
[0,150)	1.738857e-05	1.383661e-05	1.633080e-02
[150,205)	9.882397e-01	1.201782e-11	5.492948e-03
[205,270)	1.000000e+00	1.337070e-06	1.060467e-03
[270,312)	1.290321e-01	1.815878e-05	1.780064e-02
[312,363)	9.999918e-01	5.725818e-07	3.661433e-02
[363,385]	7.040308e-01	4.579072e-06	3.334993e-03

<https://doi.org/10.1371/journal.pone.0285624.t001>



**Fig 4. Susceptible and exposed estimation on the interval  $I_1 = [0, 150)$ , according to the model 4 with  $\hat{\delta}_1^2 = \sum_{j=1}^n (I(t_j) - \phi_3(t_j))^2 / n$ .**

<https://doi.org/10.1371/journal.pone.0285624.g004>



**Fig 5. Negative log-likelihood function to  $\beta \in [0, 1]$ ,  $\nu = 1.201782 \times 10^{-11}$  and  $\gamma = 5.492948 \times 10^{-3}$  on the interval  $[150, 205)$  for COVID-19 data from Bogota D.C.**

<https://doi.org/10.1371/journal.pone.0285624.g005>

Note in the Fig 3 that on the interval  $[312, 363)$ , the recovered population is over-estimated. Other people could be deficiently estimated as is observed in Fig 4, whose graph is on the interval  $I_1 = [1, 150)$  we have that the exposed population is a maximum 10,410,375, which is not valid since the people of Bogota and its metropolitan area is less than 9,000,000 (see [34]).

On the other hand, when we graph the negative log-likelihood function on the interval  $[150, 205)$  for  $\beta$  with  $\beta \in [0, 1]$ ,  $\nu = 1.201782 \times 10^{-11}$  and  $\gamma = 5.492948 \times 10^{-3}$  (Fig 5) note that this function does not have a monotone behavior. Therefore, if an initial value of  $\beta$  is near to a local minimum when the `mle2` function is used since estimation can vary significantly. The estimation of  $\beta$  is necessary for the susceptible and exposed population data using the approximation of the Euler method as given in the Eq 8.

### Data modeling

Since we have recovered data for COVID-19 data from Bogota, we initially propose an estimation method based on that data re-writing the model 2, taking the notation given by  $\phi_k(t_j)$ ,  $A_k(t_j)$ , as

$$\text{Model 2: } A_k(t) = \phi_k(t) + \epsilon_k, \epsilon_k \sim N(0, \delta_2^2). \tag{5}$$

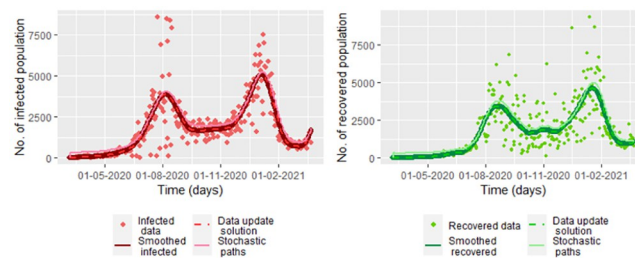


where  $A_k(t_j) \sim N(\phi_k(t_j), \hat{\delta}_2^2), j = 1, \dots, n$  with

$$\hat{\delta}_2^2 = \sum_{i=1}^n \frac{(I(t_i) - \phi_3(t_i))^2}{n} + \sum_{i=1}^n \frac{(R(t_i) - \phi_4(t_i))^2}{n}. \tag{6}$$

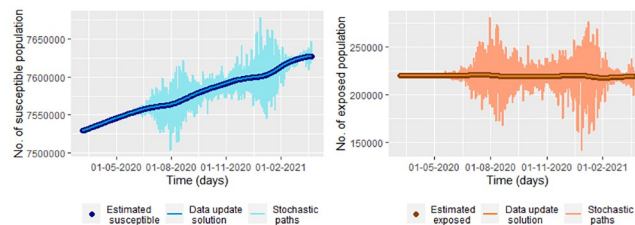
Note that in the model 5, we have noise on the infected and recovered data, which is all the available data. We do not consider the possible correlation between the infected and recovered population to simplify the calculus. We observed the inflection points graphically for the recovered population, which can be seen in Fig 6. Based on these points, combined with the points with changes in control measures on the infected population regarding [4], we chose eight different intervals for which the parameters change, under the influence of the recovered data. In the Fig 6, we also can see the estimations of the model 5 with  $\hat{\delta}_2^2$  given by the Eq 6. Note that the difference between Figs 4 and 7 is that the first one shows the estimation under the loss function uniquely considering the infected data (Model 1), while the second one shows the estimation under the loss function considering the infected data and the recovered data (Model 2). In this way, taking a loss function involving the two types of data improves the estimation of the recovered population.

As we could note in the Fig 4, when we use the estimated model 5 with  $\hat{\delta}_2^2$  given by 6, the susceptible and exposed population can be over-estimated or under-estimated. According to the estimations (see Fig 7), for day 150, there are 12,535,024 of the exposed population, which is not valid. It also can be seen that on day 150, the susceptible population is 38 approximately, something illogical since around day 300 there are more than 5,000 reported cases (see Fig 2).



**Fig 6.** Estimated infected and recovered population of the 5 model with  $\hat{\delta}_2^2$  given by the Eq 6 for each considered interval  $I_s, s = 1, \dots, 8$ . We use the m1e2 function (parameters on the Table 2).

<https://doi.org/10.1371/journal.pone.0285624.g006>



**Fig 7.** Susceptible and exposed estimation on the interval  $I_1 = [1, 150]$  according to the model 5 with  $\hat{\delta}_2^2$  given by the Eq 6.

<https://doi.org/10.1371/journal.pone.0285624.g007>

**Table 2. Estimated parameters of the 5 model with  $\hat{\delta}_2^2$  given by 6, for each considered interval  $I_s, s = 1, \dots, 8$ . We use the `m1e2` function.**

Period of time ( $\tau_s$ )	$\beta_{\tau_s}$	$\nu_{\tau_s}$	$\gamma_{\tau_s}$
[0,150)	9.999e-01	2.523e-06	8.115e-03
[150,177)	5.415e-01	3.301e-06	1.500e-02
[177,205)	1.000e+00	7.454e-29	4.098e-19
[205,270)	1.000e+00	1.337e-06	1.060e-03
[270,312)	1.290e-01	1.816e-05	1.780e-02
[312,330)	9.817e-01	1.941e-15	1.039e-02
[330,363)	1.000e+00	1.000e-06	3.806e-19
[363,385]	7.040e-01	4.579e-06	3.335e-03

<https://doi.org/10.1371/journal.pone.0285624.t002>

Figs 4 and 7 made us think about studying an estimation method that also considers the susceptible and exposed population. For that, we propose initially a parametric method of estimation where first the exposed and susceptible population is estimated. After, we find the necessity of establishing a data update estimation, which we propose in the following subsection.

### Estimation of the SEIR model from real data

This section will propose an updated data method that searches for a good fit to avoid overestimating parameters such as recovery rate  $\gamma$ . It is not adequate to take only one interval since there are some different conditions of infection spread according to local government [5]. The model to be estimated by a theoretical approach is the Model 2 of the Eq 2.

### Parameter estimation on the SEIR model with susceptible and exposed missing data

We have the data corresponding to infected and recovered population per day from Bogota (Colombia), whereby the other data (susceptible and exposed population) are missing data. In this way, we will estimate the parameters using only the available data. If we think of the ordinary least squares method, we must minimize the equation

$$Q = \sum_{i=0}^n \left( I(t_i) - \hat{I}(t_i) \right)^2 + \left( R(t_i) - \hat{R}(t_i) \right)^2, \tag{7}$$

where  $\hat{I}$  and  $\hat{R}$  are the solutions of the deterministic system 1. However, we do not have the analytical solution for  $I$  and  $R$  of the system 1, the reason why  $\hat{I}$  and  $\hat{R}$  are the approximations given by the Euler method (see [36]),

$$\begin{aligned} \hat{S}(t_j) &= S(t_{j-1}) + [\Lambda - \beta S(t_{j-1})I(t_{j-1}) - \mu S(t_{j-1})](t_j - t_{j-1}) \\ \hat{E}(t_j) &= E(t_{j-1}) + [\beta S(t_{j-1})I(t_{j-1}) - \nu E(t_{j-1}) - \mu E(t_{j-1})](t_j - t_{j-1}) \\ \hat{I}(t_j) &= I(t_{j-1}) + [\nu E(t_{j-1}) - (\mu + \gamma)I(t_{j-1})](t_j - t_{j-1}) \\ \hat{R}(t_j) &= R(t_{j-1}) + [\gamma I(t_{j-1}) - \mu R(t_{j-1})](t_j - t_{j-1}). \end{aligned} \tag{8}$$



However, note in the Eq 8 that  $\hat{I}(t_j)$  needs the values of  $E(t_{j-1})$  reason why we have to estimate  $S(t_j)$  and  $E(t_j)$ . For it, first note that

$$S(t_j) + E(t_j) = N(t_j) - I(t_j) - R(t_j), \tag{9}$$

where  $N(t_j)$  is the total population in the time  $t_j$ .

Note that,

$$\hat{S}(t_j) + \hat{E}(t_j) = S(t_{j-1}) + E(t_{j-1}) + [\Lambda - \mu(S(t_{j-1}) + E(t_{j-1})) - vE(t_{j-1})](t_j - t_{j-1}) \tag{10}$$

Replacing  $S(t_{j-1}) + E(t_{j-1})$  by  $N(t_{j-1}) - I(t_{j-1}) - R(t_{j-1})$  and  $\hat{S}(t_j) + \hat{E}(t_j)$  by  $N(t_j) - I(t_j) - R(t_j)$ , to observe for  $E(t_{j-1})$  and  $S(t_{j-1})$  given by

$$\begin{aligned} \check{E}(t_{j-1}) &= \frac{N(t_{j-1}) - N(t_j) - I(t_{j-1}) + I(t_j) - R(t_{j-1}) + R(t_j)}{v(t_j - t_{j-1})} + \frac{\Lambda}{v} \\ &\quad - \frac{\mu}{v}(N(t_{j-1}) - I(t_{j-1}) - R(t_{j-1})) \end{aligned} \tag{11}$$

$$\check{S}(t_{j-1}) = N(t_{j-1}) - \check{E}(t_{j-1}) - I(t_{j-1}) - R(t_{j-1}).$$

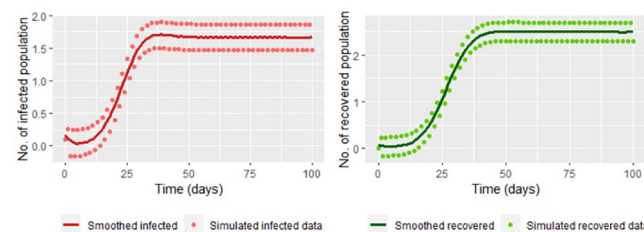
We have the estimators of  $E(t_j)$  and  $S(t_j)$  are given respectively by  $\check{E}(t_j)$  and  $\check{S}(t_j)$ . For observing that the estimators let fit the data to the model we simulate data based on the SEIR model:

1. Using the function `ode` of the library `deSolve` in R we generate the solution from 1 to 100 for the parameters  $\Lambda = 4, \beta = 0.2, v = 0.1, \gamma = 0.3$  and  $\mu = 0.2$  and initial value  $(4, 0, 0.1, 0)$ . These values we note them as  $S(t_i), E(t_i), R(t_i)$  with  $t_0 = 0, \dots, t_{100} = 100$
2. Adding 0.2 to  $I(t_i)$  and  $R(t_i)$  with  $i = 1, 3, \dots, 99$ . Subtracting 0.2 to  $I(t_i)$  and  $R(t_i)$  with  $i = 0, 2, \dots, 100$ .

In the Fig 8 we graph the simulated data in the steps 1 and 2 join with the smoothed functions using `frfast` in R.

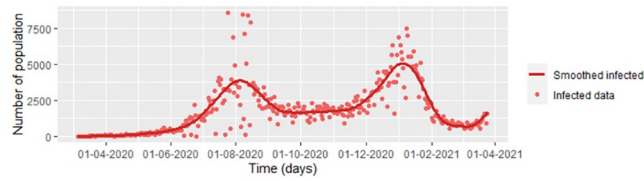
Now, we calculate  $\check{S}(t_j)$  and  $\check{E}(t_j)$  (Eq 11) with  $t_j = 1, \dots, 100$  for comparing the estimation with the “real” data. Fig 9 shows that comparison.

The susceptible and exposed population is over or underestimated when taking the data directly. We use smoothing techniques to reduce the effect of variations for the infected and recovered data (for example, we use `frfast` in R) as it can be seen in the Fig 10



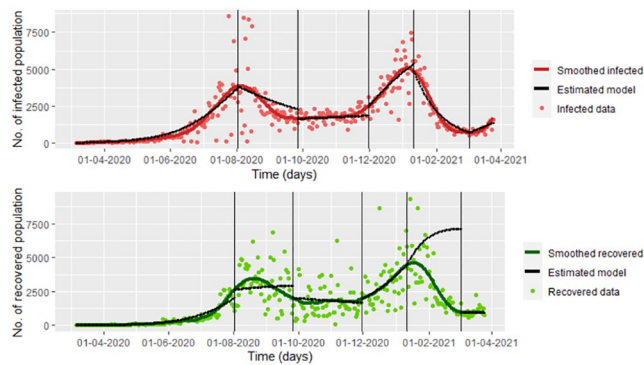
**Fig 8. Infected and recovered simulated date built in the steps 1 and 2 and their smoothing.**

<https://doi.org/10.1371/journal.pone.0285624.g008>



**Fig 9. Susceptible and exposed population estimated by  $\check{S}(t_j)$  and  $\check{E}(t_j)$  (Eq 11) without smoothing infected and recovered data versus solutions for susceptible and recovered of the step 1 (modeled exposed and recovered, respectively).**

<https://doi.org/10.1371/journal.pone.0285624.g009>



**Fig 10. Susceptible and exposed population estimated by  $\check{S}(t_j)$  and  $\check{E}(t_j)$  (Eq 11) versus solutions for susceptible and recovered of the step 1 (modeled exposed and recovered, respectively).**

<https://doi.org/10.1371/journal.pone.0285624.g010>

Thus, we suggest smoothing the data to estimate the susceptible and exposed population better. On the other hand, we also can have a good estimation by smoothing the values of  $\check{S}(t_j)$  and  $\check{E}(t_j)$ . Note in the Eq 11 we have assumed  $\nu$  as the known parameter. Therefore we would have to estimate  $\beta$  and  $\gamma$  using the data. We consider the data given by  $I(t_j)$ ,  $R(t_j)$  and  $S(t_j) \approx \check{S}(t_j)$ ,  $E(t_j) \approx \check{E}(t_j)$ . Thus we can apply the least-squares method by minimizing the equation

$$U = \sum_{i=0}^n \left( (\check{S}(t_i) - \hat{S}(t_i))^2 + (\check{E}(t_i) - \hat{E}(t_i))^2 + (I(t_i) - \hat{I}(t_i))^2 + (R(t_i) - \hat{R}(t_i))^2 \right), \quad (12)$$

taking taking  $\Delta t = t_j - t_{j-1}$  for all  $j = 1, \dots, n$ , we have that  $U$  is given by

$$\begin{aligned} U = & \sum_{j=0}^{n+1} (\check{S}(t_j) - \check{S}(t_{j-1}) - \Lambda \Delta t + \beta \check{S}(t_{j-1}) I(t_{j-1}) \Delta t + \mu \check{S}(t_{j-1}) \Delta t)^2 + \\ & \sum_{j=0}^{n+1} (\check{E}(t_j) - \check{E}(t_{j-1}) - \beta \check{S}(t_{j-1}) I(t_{j-1}) \Delta t + \nu \check{E}(t_{j-1}) \Delta t + \mu \check{E}(t_{j-1}) \Delta t)^2 + \\ & \sum_{j=1}^{n+1} (I(t_j) - I(t_{j-1}) - \nu \check{E}(t_{j-1}) \Delta t + (\mu + \gamma) I(t_{j-1}) \Delta t)^2 + \\ & \sum_{j=1}^{n+1} (R(t_j) - R(t_{j-1}) - \gamma I(t_{j-1}) \Delta t + \mu R(t_{j-1}) \Delta t)^2. \end{aligned} \quad (13)$$

We can find the value of  $\hat{\beta}$  by minimizing  $U$  as follows

$$\begin{aligned} \frac{\partial U}{\partial \beta} = & 2 \sum_{j=1}^{n+1} (\check{S}(t_j) - \check{S}(t_{j-1}) - \Lambda \Delta t + \beta \check{S}(t_{j-1}) I(t_{j-1}) \Delta t + \mu \check{S}(t_{j-1}) \Delta t) \\ & \cdot \check{S}(t_{j-1}) I(t_{j-1}) \Delta t + 2 \sum_{j=1}^{n+1} (\check{E}(t_j) - \check{E}(t_{j-1}) - \beta \check{S}(t_{j-1}) I(t_{j-1}) \Delta t \\ & + v \check{E}(t_{j-1}) \Delta t + \mu \check{E}(t_{j-1}) \Delta t) \cdot (-\check{S}(t_{j-1}) I(t_{j-1}) \Delta t) = 0. \end{aligned} \tag{14}$$

Obtaining that  $\hat{\beta}$  is given by

$$\hat{\beta} = \frac{c}{2 \Delta t \sum_{j=1}^{n+1} \check{S}^2(t_{j-1}) I^2(t_{j-1})}, \tag{15}$$

where

$$\begin{aligned} c = & (1 - \mu \Delta t) \sum_{j=1}^{n+1} \check{S}^2(t_{j-1}) I(t_{j-1}) - \sum_{j=1}^{n+1} \check{S}(t_j) \check{S}(t_{j-1}) I(t_{j-1}) \\ & + \Lambda \Delta t \sum_{j=1}^{n+1} \check{S}(t_{j-1}) I(t_{j-1}) - (1 - \hat{v} \Delta t - \mu \Delta t) \sum_{j=1}^{n+1} \check{E}(t_{j-1}) \check{S}(t_{j-1}) I(t_{j-1}) \\ & + \sum_{j=1}^{n+1} \check{E}(t_j) \check{S}(t_{j-1}) I(t_{j-1}). \end{aligned}$$

On the other hand, for  $\gamma$  we have

$$\begin{aligned} \frac{\partial U}{\partial \gamma} = & 2 \sum_{j=1}^{n+1} (I(t_j) - I(t_{j-1}) - v E(t_{j-1}) \Delta t + (\mu + \gamma) I(t_{j-1}) \Delta t) I(t_{j-1}) \Delta t + \\ & 2 \sum_{j=1}^{n+1} (R(t_j) - R(t_{j-1}) - \gamma I(t_{j-1}) \Delta t + \mu R(t_{j-1}) \Delta t) (-I(t_{j-1}) \Delta t) = 0. \end{aligned} \tag{16}$$

Obtaining that  $\hat{\gamma}$  is given by

$$\hat{\gamma} = \frac{a + v \sum_{i=1}^{n+1} \check{E}(t_{j-1}) I(t_{j-1})}{2 \sum_{i=1}^{n+1} I^2(t_{j-1})}, \tag{17}$$

where

$$\begin{aligned} a = & \frac{(1 - \mu \Delta t)}{\Delta t} \left( \sum_{i=1}^{n+1} I^2(t_{j-1}) - \sum_{i=1}^{n+1} R(t_{j-1}) I(t_{j-1}) \right) - \frac{1}{\Delta t} \sum_{i=1}^{n+1} I(t_i) I(t_{j-1}) \\ & + \frac{1}{\Delta t} \sum_{i=1}^{n+1} R(t_i) I(t_{j-1}). \end{aligned}$$

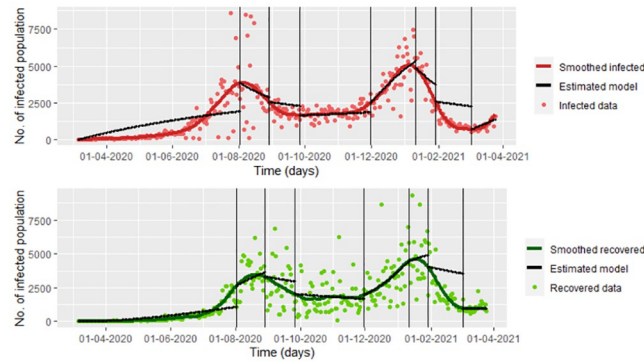
Note that the estimators are minimum due to the Hessian matrix is given by

$$\mathcal{H}_{\beta,\gamma} = \begin{pmatrix} 4 \sum_{i=1}^{n+1} \check{S}^2(t_{j-1}) I^2(t_{j-1}) \Delta^2 t & 0 \\ 0 & 4 \sum_{i=1}^{n+1} I^2(t_{j-1}) \Delta^2 t \end{pmatrix}$$

such as  $4 \sum_{i=1}^{n+1} \check{S}^2(t_{j-1}) I^2(t_{j-1}) \Delta^2 t > 0$  and

$$\det(\mathcal{H}_{\beta,\gamma}) = 16 \Delta^4 t \left( \sum_{i=1}^{n+1} \check{S}^2(t_{j-1}) I^2(t_{j-1}) \right) \left( \sum_{i=1}^{n+1} I^2(t_{j-1}) \right) > 0,$$

therefore  $(\hat{\beta}, \hat{\gamma})$  is a minimum, being these the least squares estimators. In the Fig 11 we graph the solution of the system 1 (using the ode function) taking  $\Lambda = 4, v = 0.1, \mu = 0.2, \beta \approx \hat{\beta}$  and  $\gamma \approx \hat{\gamma}$ ; where  $\hat{\beta}$  and  $\hat{\gamma}$  are calculated without smoothing the infected ( $I(t_j)$ ) and recovered ( $R$



**Fig 11. Solution of the system 1 taking  $\Lambda = 4, v = 0.1, \mu = 0.2, \beta = \hat{\beta}$  and  $\gamma = \hat{\gamma}$ ; where  $\hat{\beta}$  and  $\hat{\gamma}$  are calculated without smoothing the infected and recovered data simulated (Solution 1). On the other hand, the real solution, which is the solution of the system 1 taking  $\Lambda = 4, v = 0.1, \mu = 0.2, \beta = 0.2$  and  $\gamma = 0.3$  (Solution 2) is also graphed.**

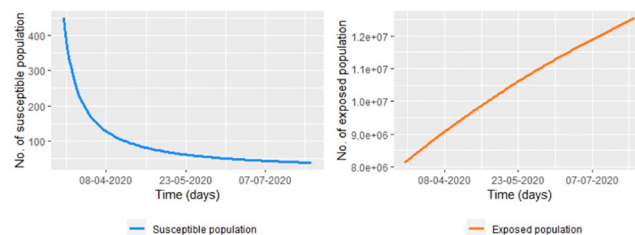
<https://doi.org/10.1371/journal.pone.0285624.g011>

( $t_j$ ) data simulated in the steps 1 and 2, and whose values are  $\hat{\beta} = 0.5581449$  and  $\hat{\gamma} = 0.2645688$ . On the same graph, we show the solutions according to ode function with  $\Lambda = 4, v = 0.1, \mu = 0.2, \beta = 0.2$  and  $\gamma = 0.3$ .

Note that we don't have a good estimation for the infected and recovered data without smoothing. It also can be seen that the solution for the susceptible and recovered is far from the corresponding smoothed population estimated. This can be corrected by smoothing the infected and recovered data (in our case, the data simulated by the steps 1 and 2) before calculating  $\hat{\beta}$  and  $\hat{\gamma}$ , whose values are  $\hat{\beta} = 0.199841$  and  $\hat{\gamma} = 0.2934448$  close to  $\beta$  and  $\gamma$ , respectively. Note in Fig 12 that smoothing the data lets a good fit since smoothed data is near to the solution of the system.

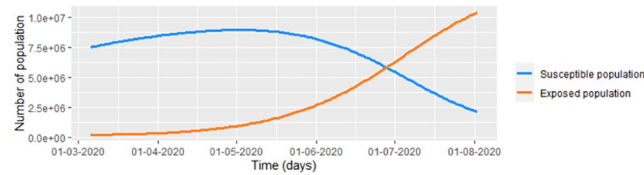
Following the last paragraph, we taking the smoothed infected and recovered data of COVID-19 from Bogota (Fig 2) to estimate the susceptible and exposed population under the pandemic (Fig 13). Calculating  $\check{S}(t_j)$  and  $\check{E}(t_j)$  (Eq 11) with  $\Lambda = (9.42/1000)\bar{p} \approx 73,660, \mu = 4/1000$  and the incubation rate of COVID-19 given by  $v = 1/5.2$  ( $1/v = 5.2$  days is the incubation time according to [37]).

Note that the susceptible population always increases. This is why  $\Lambda \gg \mu \max_i(p_i)$  with  $p_i$  the projected population from Bogota in the day  $i$ . The reason why there is such a large increase in the population from Bogota is the migration mostly of people from Venezuela regarding to Migración Colombia [38]. On the other hand, the exposed population presents



**Fig 12. Solution of the system 1 taking  $\Lambda = 4, v = 0.1, \mu = 0.2, \beta = \hat{\beta}$  and  $\gamma = \hat{\gamma}$ ; where  $\hat{\beta}$  and  $\hat{\gamma}$  are calculated by smoothing the infected and recovered data simulated (Solution 1). On the other hand, the real solution, which is the solution of the system 1 taking  $\Lambda = 4, v = 0.1, \mu = 0.2, \beta = 0.2$  and  $\gamma = 0.3$  (Solution 2) is also graphed.**

<https://doi.org/10.1371/journal.pone.0285624.g012>



**Fig 13. Projected population (according to DANE) from Bogota during the first 385 days of COVID-19 in Bogota. Susceptible and exposed estimations using the Eq 11.**

<https://doi.org/10.1371/journal.pone.0285624.g013>

peaks and valleys which match with the peaks and valleys of the number of infected in Bogota. In the Table 3 we estimate the parameters of the eight intervals under which we consider different infection conditions. Note that the recovered rate could be over-estimated since the values of  $\hat{\gamma}$  are more than 1, which indicates that the recuperation time of COVID-19 is less than one day which is false [39].

Despite using a least-squares method for estimating the parameters and taking some intervals with similar infection conditions, the Table 3 suggests using another estimation method for modeling the COVID-19 with the SEIR model. For this reason, we propose an updated data estimation in the next subsection.

### Parameter estimation on the SEIR model with susceptible, exposed and recovered missing data

In this case, we do not have data corresponding to susceptible, exposed, and recovered populations each time. We could only apply the least squares on the infected population.

$$S = \sum_{j=0}^n (I(t_j) - \hat{I}(t_j))^2. \tag{18}$$

However, we would need the values of  $E(t_{j-1})$  according to the Eq 8, from where we can see that

$$\begin{aligned} \hat{S}(t_j) + \hat{E}(t_j) + \hat{R}(t_j) = & S(t_{j-1}) + E(t_{j-1}) + R(t_{j-1}) + [\Lambda - \mu(S(t_{j-1}) + E(t_{j-1}) \\ & + R(t_{j-1})) - \nu E(t_{j-1}) + \gamma I(t_{j-1})](t_j - t_{j-1}) \end{aligned} \tag{19}$$

**Table 3. Parameter estimation using  $\hat{\beta}$  and  $\hat{\gamma}$  using the Eqs 15 and 17 for each considered interval  $I_s, s = 1, \dots, 8$ , respectively. We take  $\check{S}(t_j)$  and  $\check{E}(t_j)$  according to the Fig 13.**

Period of time ( $\tau_s$ )	$\hat{\beta}_{\tau_s}$	$\hat{\gamma}_{\tau_s}$
[1,150)	2.353e-06	8.461e+00
[150,177)	1.600e-06	5.767e+00
[177,205)	2.829e-06	1.019e+01
[205,270)	2.997e-06	1.082e+01
[270,312)	1.309e-06	4.736e+00
[312,330)	1.375e-06	4.981e+00
[330,363)	3.733e-06	1.350e+01
[363,385)	5.105e-06	1.850e+01

<https://doi.org/10.1371/journal.pone.0285624.t003>

Replacing  $S(t_{j-1}) + E(t_{j-1}) + R(t_{j-1})$  by  $N(t_{j-1}) - I(t_{j-1})$  and  $\hat{S}(t_j) + \hat{E}(t_j) + \hat{R}(t_j)$  by  $N(t_j) - I(t_j)$ , we obtain an estimator for  $E(t_{j-1})$

$$\begin{aligned} \check{E}(t_{j-1}) = & \frac{N(t_{j-1}) - N(t_j) - I(t_{j-1}) + I(t_j)}{v(t_j - t_{j-1})} + \frac{\Lambda}{v} \\ & - \frac{\mu}{v} [N(t_{j-1}) - I(t_{j-1}) + \gamma I(t_{j-1})] \end{aligned} \tag{20}$$

Note that we assumed  $\gamma$  and  $v$  as known parameters. In such manner, we can estimate  $R(t_{j-1})$  by  $\hat{R}(t_{j-1})$  (Eq 8) taking  $R(t_0) = 0$  and  $S(t_{j-1})$  by

$\check{S}(t_{j-1}) = N(t_{j-1}) - \check{E}(t_{j-1}) - I(t_{j-1}) - \hat{R}(t_{j-1})$ . We want to minimize for finding an estimator for  $\beta$

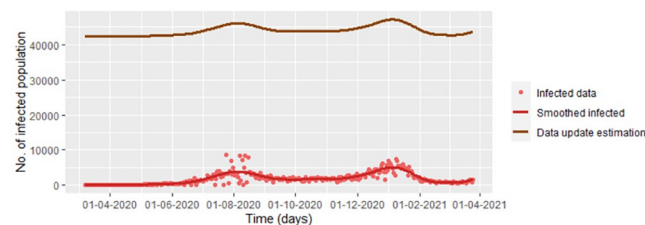
$$V = \sum_{j=0}^n ((\check{S}(t_j) - \hat{S}(t_j))^2 + (\check{E}(t_j) - \hat{E}(t_j))^2 + (I(t_j) - \hat{I}(t_j))^2 + (R(t_j) - \hat{R}(t_j))^2), \tag{21}$$

where  $\check{R}(t_i) = \hat{R}(t_i)$ . Actually, independently if we have  $\check{R}(t_i) \neq \hat{R}(t_i)$ , for minimizing  $V$  we have that the derivative respect to  $\beta$  of  $\sum_{i=0}^n (\hat{R}(t_i) - \check{R}(t_i))^2$  is equal to 0 since  $\beta$  does not appear at  $\hat{R}(t_i)$ .

Analogously, following what we did for minimizing  $U$  on the Eq 12 we have that  $\hat{\beta}$  is given by the Eq 15, which is a minimum due to  $\partial^2 V / \partial \beta^2 = 4 \sum_{i=1}^{n+1} \check{S}^2(t_{j-1}) I^2(t_{j-1}) > 0$ .

In Fig 14, we graph the estimations for the susceptible ( $\check{S}(t_j)$ ), exposed ( $\check{E}(t_j)$ ) and recovered ( $\hat{R}(t_j)$ ) population, from the infected data simulated in the steps 1 and 2 (considering unknown recovered data). In addition, we graph the solution of the system taking the parameters as the step 1, except for  $\beta$ , which is taken by  $\beta \approx \hat{\beta} = 0.1397632$  calculated of the Eq 15, under  $\check{S}(t_j) \approx \check{S}(t_j)$ ,  $\check{E}(t_j) \approx \check{E}(t_j)$  and the infected data without smoothing.

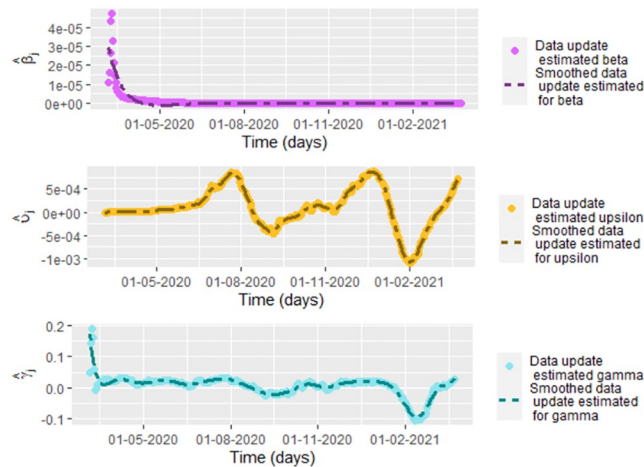
If we take the estimations for the susceptible ( $\check{S}(t_j)$ ), exposed ( $\check{E}(t_j)$ ) and recovered ( $\hat{R}(t_j)$ ) population, calculated only from the smoothed infected data (from the infected data simulated in the steps 1 and 2) using the function `frfast` from the package `npregfast` in R. We also calculate  $\hat{\beta}$  using  $\check{S}(t_j) \approx \check{S}(t_j)$ ,  $\check{E}(t_j) \approx \check{E}(t_j)$  (using the smoothed simulated data) and the smoothed infected data, and so  $\beta \approx \hat{\beta} = 0.1397632$ , which is too far from  $\beta = 0.2$ . That implies that not necessarily smooth the data drives to a good fit (see Fig 15).



**Fig 14. Solution of the system 1 taking  $\Lambda = 4$ ,  $v = 0.1$ ,  $\mu = 0.2$ ,  $\gamma = 0.3$  and  $\beta = \hat{\beta}$  (Solution 1).**  $\hat{\beta}$  is calculated using  $\check{S}(t_j), \check{E}(t_j)$  (based on the smoothed  $I(t_j)$ ) and the infected data without smoothing. On the other hand, the real solution, which is the solution of the system 1 taking  $\Lambda = 4$ ,  $v = 0.1$ ,  $\mu = 0.2$ ,  $\beta = 0.2$  and  $\gamma = 0.3$  (Solution 2) is also graphed.

<https://doi.org/10.1371/journal.pone.0285624.g014>





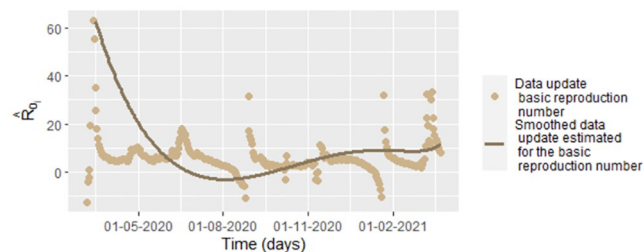
**Fig 15. Solution of the system 1 taking  $\Lambda = 4, \nu = 0.1, \mu = 0.2, \gamma = 0.3$  and  $\beta = \hat{\beta}$  (Solution 1).**  $\hat{\beta}$  is calculated using  $\hat{S}(t_j), \hat{E}(t_j)$  (based on the smoothed  $I(t_j)$ ) and the smoothed infected data. On the other hand, the real solution, which is the solution of the system 1 taking  $\Lambda = 4, \nu = 0.1, \mu = 0.2, \beta = 0.2$  and  $\gamma = 0.3$  (Solution 2) is also graphed.

<https://doi.org/10.1371/journal.pone.0285624.g015>

Note in the bottom right of the Fig 15 that there is a good estimation for the recovered population, being too close  $\hat{R}(t_j)$  to  $R(t_j)$ . We can estimate the susceptible and exposed population using the Eq 22.

$$\begin{aligned} \overline{E}(t_{j-1}) &= \frac{N(t_{j-1}) - N(t_j) - I(t_{j-1}) + I(t_j) - \hat{R}(t_{j-1}) + \hat{R}(t_j)}{\nu(t_j - t_{j-1})} + \frac{\Lambda}{\nu} \\ &\quad - \frac{\mu}{\nu}(N(t_{j-1}) - I(t_{j-1}) - \hat{R}(t_{j-1})) \\ \overline{S}(t_{j-1}) &= N(t_{j-1}) - \overline{E}(t_{j-1}) - I(t_{j-1}) - \hat{R}(t_{j-1}). \end{aligned} \tag{22}$$

Estimating the susceptible, exposed and recovered population respectively by  $\overline{S}(t_j), \overline{R}(t_j)$  and  $\hat{R}(t_j)$ , note in the Fig 16 that these populations are being good estimated.



**Fig 16. Solution of the system 1 taking  $\Lambda = 4, \nu = 0.1, \mu = 0.2, \gamma = 0.3$  and  $\beta = \hat{\beta}$  (Solution 1).**  $\hat{\beta}$  is calculated using  $\hat{S}(t_j), \hat{E}(t_j)$  (based on the smoothed  $I(t_j)$ ) and the smoothed infected data. On the other hand, the real solution, which is the solution of the system 1 taking  $\Lambda = 4, \nu = 0.1, \mu = 0.2, \beta = 0.2$  and  $\gamma = 0.3$  (Solution 2) is also graphed.

<https://doi.org/10.1371/journal.pone.0285624.g016>

Calculating  $\hat{\beta}$  (Eq 15) based on  $\overline{S}(t_j)$ ,  $\overline{R}(t_j)$  and the smoothed infected data, we have  $\hat{\beta} = 0.2002431$ , which is close to  $\beta = 0.2$  which with the Fig 16 indicate a good estimation.

### A data update approach

Exogenous variables such as vaccination and quarantines interact with the infections, and we did not consider them in the present model. However, the SEIR model is one of the simplest to describe an epidemic’s behavior. For this reason, we propose a data update method of parameter estimation based on previous approximations to the unknown populations who are susceptible and exposed. This method estimates the parameters for each time  $t_j$ , based on the available data of the previous day. Its implementation is different for each model, and it depends on the previous known parameters. This focus was approached by [40], under a regression Poisson model applied only on age groups, to predict the number of cases and deaths of COVID-19 in Italy, taking as regression variable the time. [41] estimate and map the prevalence of Chagas disease among adults in the United States, based on some small population subgroups at the public micro-area (PUMA) level for mapping. [42] uses Markov Chain Monte Carlo (MCMC) method to fit an SEIR-type model to the data of the cumulative number of laboratory-confirmed 2019-nCov cases from the National Health Commission of the People’s Republic of China. Finally, [43] suggests using a bayesian model based on Newtonian equations of an ordinary differential system to estimate the parameters. In this paper, we have given a general estimation of the parameters. However, it is not used with real data. In our paper, we propose a novel method whose purpose is estimating each parameter on time  $t_j$  from infected and recovered data of the previous day, that is  $I(t_{j-1})$  and  $R(t_{j-1})$ . We are going to consider  $\hat{S}(t_j) \approx \check{S}(t_j)$ ,  $\hat{E}(t_j) \approx \check{E}(t_j)$ ,  $\hat{I}(t_j) \approx I(t_j)$  and  $\hat{R}(t_j) \approx R(t_j)$  to estimate the parameters based on a data update for  $\beta$  and  $\gamma$ . Solving for  $\gamma$  on the Eq 8 we have

$$\hat{\gamma}_j = \frac{R(t_j) - R(t_{j-1}) + \mu R(t_{j-1})\Delta t}{I(t_{j-1})\Delta t}. \tag{23}$$

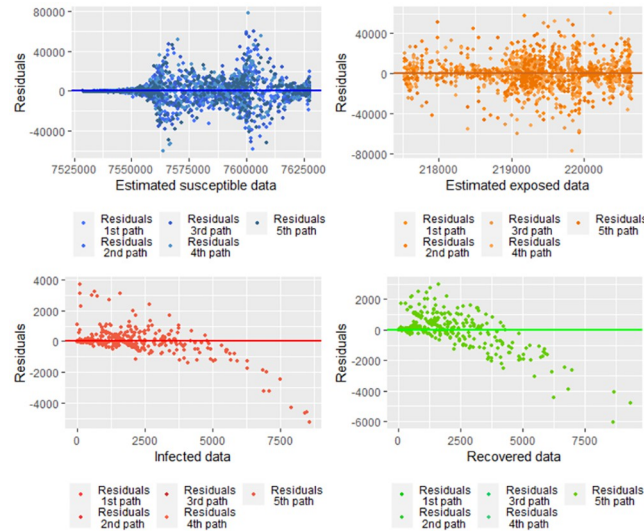
In the Eq 23, we fit the recovered data using the data update approach. On the other hand, we solve  $\beta$  from the Eq 8 as follows

$$\tilde{\beta}_j = \frac{\check{S}(t_{j-1}) - \check{S}(t_j) + \Lambda\Delta t - \mu\check{S}(t_{j-1})\Delta t}{\check{S}(t_{j-1})I(t_{j-1})\Delta t}. \tag{24}$$

In the Eq 24, we fit the susceptible estimated data using the data update approach. Eq 25 is based completely in the Eq 8, which was obtained of replacing  $\beta$  and  $\gamma$  by  $\tilde{\beta}_j$  and  $\hat{\gamma}_j$ , respectively. In Fig 17, we graph the values given by 25 with the infected data and its smoothing.

$$\begin{aligned} \check{S}^*(t_j) &= \check{S}(t_{j-1}) + [\Lambda - \tilde{\beta}_j\check{S}(t_{j-1})I(t_{j-1}) - \mu\check{S}(t_{j-1})](t_j - t_{j-1}) \\ \check{E}^*(t_j) &= \check{E}(t_{j-1}) + [\tilde{\beta}_j\check{S}(t_{j-1})I(t_{j-1}) - v\check{E}(t_{j-1}) - \mu\check{E}(t_{j-1})](t_j - t_{j-1}) \\ \check{I}^*(t_j) &= I(t_{j-1}) + [v\check{S}(t_{j-1}) - (\mu + \hat{\gamma}_j)I(t_{j-1})](t_j - t_{j-1}) \\ \check{R}^*(t_j) &= R(t_{j-1}) + [\hat{\gamma}_jI(t_{j-1}) - \mu R(t_{j-1})](t_j - t_{j-1}), \end{aligned} \tag{25}$$

However, when we compare with the COVID-19 data for Bogota, we realize that this does not work for the infected population for which there is an overestimation, as can be seen in Fig



**Fig 17. Estimated infected population under the Eq 25 for smoothed COVID-19 Bogota data using the estimators given by  $v = 1/5.2$ ,  $\hat{\beta}_j$  and  $\hat{\gamma}_j$  24 and 23, respectively.**

<https://doi.org/10.1371/journal.pone.0285624.g017>

17. Reason why there is an over-estimation is that  $\check{S}(t_{j-1}) - \check{S}(t_j)$  is a big number in comparison with  $\Lambda \Delta t - \mu \check{S}(t_{j-1})$ . On the other hand, despite having a good estimation for the susceptible population, it is important to say that there are exogenous variables that were not considered, in particular, the vaccination which started in February 17<sup>th</sup>/2021, reason which the infected population avoids of being good estimated. However, the focus given by the equations 24 and 23 let us estimate when there were peaks and valleys, in spite of the over-estimation of the infected population.

Estimators given by 24 and 23 fit only the susceptible and the recovered population, which is why the infected and the exposed population is not well estimated. Therefore, we propose an estimation, but considering  $\beta$ ,  $v$ , and  $\gamma$  under other data update approach, for which  $\gamma$  is fitted for the recovered population,  $v$  for the infected population and  $\beta$  for the exposed population.

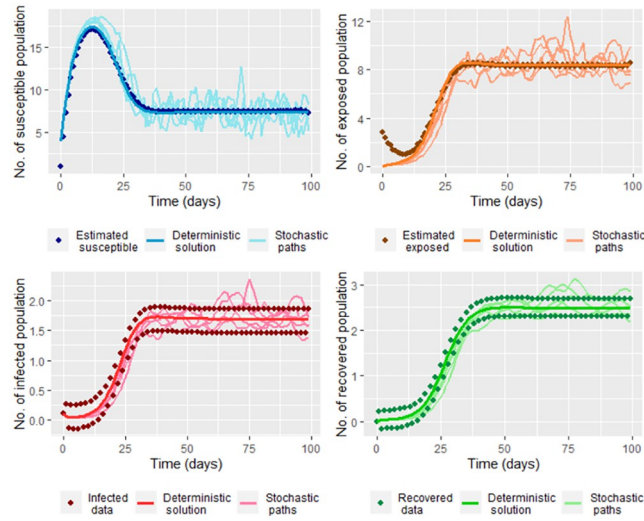
It is important to highlight that  $v$  is being re-estimated because we have to use the exposed population for the estimation. We could consider the values of  $\check{E}(t_j)$  as an indication of the behavior of the exposed population. Solving for  $\gamma$ ,  $v$  and  $\beta$  on the Eq 8 we obtain  $\hat{\gamma}_j$ ,  $\hat{v}_j$  and  $\hat{\beta}_j$  given respectively by 23,

$$\hat{v}_j = \frac{I(t_j) - I(t_{j-1}) + (\mu + \hat{\gamma}_j)I(t_{j-1})\Delta t}{\check{E}(t_{j-1})\Delta t}, \tag{26}$$

and

$$\hat{\beta}_j = \frac{\check{E}(t_j) - \check{E}(t_{j-1}) + (\hat{v}_j + \mu)\check{E}(t_{j-1})\Delta t}{\check{S}(t_{j-1})I(t_{j-1})\Delta t}. \tag{27}$$

For smoothed infected and recovered COVID-19 data from Bogota note that there is a good fitting between the estimation by the data update approach method and the data, as you can see in the Fig 18. We graph the data update estimation using the Eq 28 which is a version of 8 replacing  $\beta$ ,  $v$ ,  $\gamma$ ,  $S(t_j)$  and  $E(t_j)$  by  $\hat{\beta}_j$ ,  $\hat{v}_j$ ,  $\hat{\gamma}_j$  (from the Eqs 23, 26 and 27, respectively),  $\check{S}(t_j)$



**Fig 18. Estimated populations under the Model 3 for smoothed COVID-19 Bogota data using the estimators given by  $\hat{\beta}_j$ ,  $\hat{\nu}_j$  and  $\hat{\gamma}_j$  given by 27, 26 and 23, respectively.**

<https://doi.org/10.1371/journal.pone.0285624.g018>

and  $\check{E}(t_j)$  (from the Eq 11).

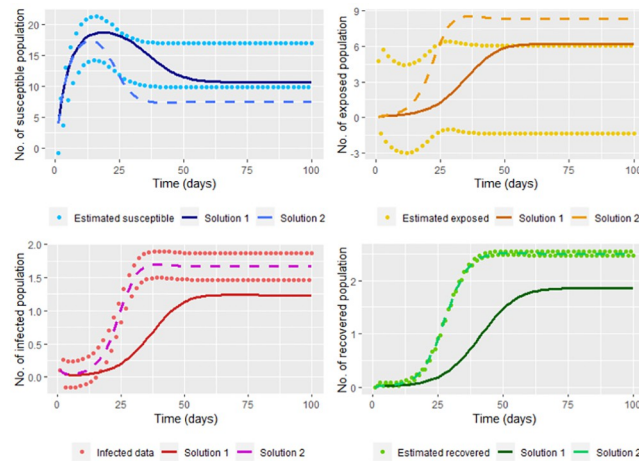
**Model 3 :**

$$\begin{aligned}
 \widehat{S}(t_j) &= \check{S}(t_{j-1}) + [\Lambda - \hat{\beta}_j \check{S}(t_{j-1}) I(t_{j-1}) - \mu \check{S}(t_{j-1})](t_j - t_{j-1}) \\
 \widehat{E}(t_j) &= \check{E}(t_{j-1}) + [\hat{\beta}_j \check{S}(t_{j-1}) I(t_{j-1}) - \hat{\nu}_j \check{E}(t_{j-1}) - \mu \check{E}(t_{j-1})](t_j - t_{j-1}) \\
 \widehat{I}(t_j) &= I(t_{j-1}) + [\hat{\nu}_j \check{S}(t_{j-1}) - (\mu + \hat{\gamma}_j) I(t_{j-1})](t_j - t_{j-1}) \\
 \widehat{R}(t_j) &= R(t_{j-1}) + [\hat{\gamma}_j I(t_{j-1}) - \mu R(t_{j-1})](t_j - t_{j-1}).
 \end{aligned}
 \tag{28}$$

Note that the model 3 given by the system 28 has a perfect fit for the indication of the exposed population. Thus, if we had the exposed data, we would have a good fit by smoothing the data.

On the other hand, the susceptible population is being over-estimated on average by 42 084 of the population. That is because we did the data update for the indication of the exposed population but not for the indication of the susceptible population. For improving the susceptible estimation we use that  $S(t_j) \approx N(t_j) - \check{E}(t_j) - \check{I}(t_j) - \check{R}(t_j)$  with  $\check{E}(t_j)$ ,  $\check{I}(t_j)$  and  $\check{R}(t_j)$  are the update data estimation of  $E(t_j)$ ,  $I(t_j)$  and  $R(t_j)$ , respectively. In the Fig 19 we see that there is a good fit between the indication susceptible data and the values of  $N(t_j) - \check{E}(t_j) - \check{I}(t_j) - \check{R}(t_j)$ . Note in the Fig 19 that the susceptible population is predicted to always increase. This due to we do not consider vaccination or other measures to control the epidemic.

For each one of the eight intervals that we take with different infection conditions of COVID-19 (Fig 6), we calculate the mean and the standard deviation for each parameter. Note in the Table 4 the means of the incubation and the recuperation rates are negative when as the number of infected as of recovered population decrease, that is, on the intervals [177, 205),



**Fig 19. A update data susceptible estimation based on  $N(t_j) - \tilde{E}(t_j) - \tilde{I}(t_j) - \tilde{R}(t_j)$  with  $\tilde{E}(t_j), \tilde{I}(t_j)$  and  $\tilde{R}(t_j)$  given by the update data method of estimation for  $E(t_j), I(t_j)$  and  $R(t_j)$ , respectively.**

<https://doi.org/10.1371/journal.pone.0285624.g019>

[312, 330) and [330, 363). For the rest of the intervals, all the parameters are positive, even when only the infected population decreases as happens on the interval [150, 177) (see Fig 6).

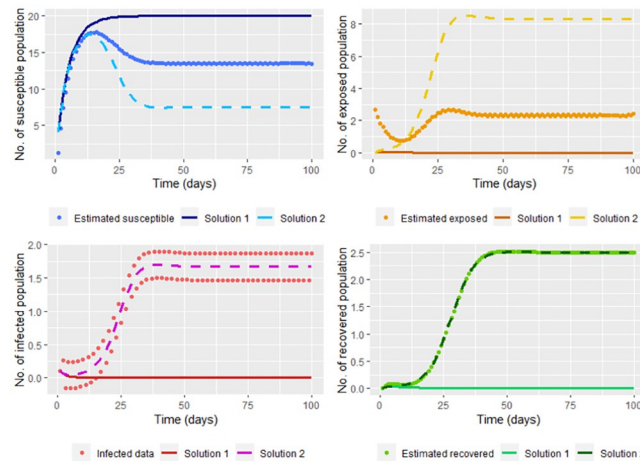
In some papers as [27–31] are considered the parameters as functions depending on the time. We can think  $\beta, v$  and  $\gamma$  as functions, whose graphs for COVID-19 data from Bogota is showed in the Fig 20. We also could interpret  $\beta, v$  and  $\gamma$  under the focus of time series, estimating the corresponding parameters using the data given by  $\beta_j, v_j$  and  $\gamma_j$  calculated respectively by 27, 26 and 23. In the Fig 20, we also graph the smoothed functions of  $\beta_j, v_j$  and  $\gamma_j$  using the function `frfast` in R.

Under the values of  $\beta_j, v_j$  and  $\gamma_j$  it is not possible to model  $\beta, v$  and  $\gamma$  as random variables since the  $\beta_j, v_j$  and  $\gamma_j$  are not independent since the parameters depend on the number on the infected and recovered people on the previous days. In the Fig 21 we graph the basic reproduction number by data update estimation,  $\hat{R}_{0j} = \frac{\hat{v}_j \hat{\beta}_j \Lambda}{\mu(\hat{v}_j + \mu)(\hat{\gamma}_j + \mu)}$ , (see [44]) for each day, where  $\hat{\beta}_j, \hat{v}_j$  and  $\hat{\gamma}_j$  given by 27, 26 and 23, respectively. In the same figure we also show the corresponding smoothing using the function `frfast` in R. Similarly that for  $\beta, v$  and  $\gamma$  estimate by data update, we also we think under this focus that  $R_0$  is a time series or a function depending on the time, but not a random variable.

**Table 4. Mean and standard deviation of the values of  $\hat{\beta}_j, \hat{v}_j$  and  $\hat{\gamma}_j$  given by 27, 26 and 23, respectively, for each period of time of the Fig 6.**

Period ( $\tau_s$ )	Mean of $\hat{\beta}_j$	SD of $\hat{\beta}_j$	Mean of $\hat{v}_j$	SD of $\hat{v}_j$	Mean of $\hat{\gamma}_j$	SD of $\hat{\gamma}_j$
[1,150)	2.382e-06	6.666e-06	2.208e-04	2.756e-04	2.313e-02	2.222e-02
[150,177)	3.276e-08	3.003e-09	4.669e-06	2.996e-04	9.426e-03	9.603e-03
[177,205)	5.816e-08	9.896e-09	-2.669e-04	1.233e-04	-1.859e-02	4.833e-03
[205,270)	6.644e-08	4.498e-09	1.230e-04	1.539e-04	2.942e-03	1.131e-02
[270,312)	3.500e-08	1.002e-08	6.441e-04	2.076e-04	1.770e-02	1.749e-03
[312,330)	2.382e-08	3.285e-09	-5.584e-04	3.532e-04	-5.225e-03	1.503e-02
[330,363)	1.062e-07	4.868e-08	-6.097e-04	3.689e-04	-7.177e-02	2.921e-02
[363,385)	1.395e-07	3.025e-08	2.774e-04	2.279e-04	7.101e-03	8.571e-03

<https://doi.org/10.1371/journal.pone.0285624.t004>

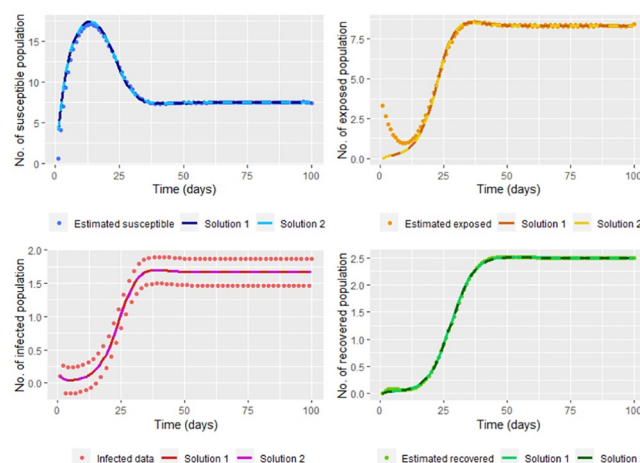


**Fig 20.** Graphs of  $\hat{\beta}_j$ ,  $\hat{\nu}_j$  and  $\hat{\gamma}_j$  given by 27, 26 and 23, respectively, with respect to the time for smoothed COVID-19 data from Bogota.

<https://doi.org/10.1371/journal.pone.0285624.g020>

In the Table 5 we calculate the mean and the standard deviation of the data update estimation of the basic reproduction number and for the corresponding smoothing. Interpreting the basic reproduction number ([45]), we conclude that the COVID-19 has been contagious high in Bogota, even when the number of infected population has decreased. The conditions for fewer infections are when the number of infected populations decreases and the recovered increases (interval [150, 177]).

- “If  $R_0 < 1$  then on average an infected individual produces less than one new infected individual throughout its infectious period, and the infection cannot grow” [45],
- “If  $R_0 > 1$  then each infected individual produces, on average, more than one new infection, and the disease can invade the population” [45],



**Fig 21.** Graphs of the estimated basic reproduction number,  $\hat{R}_{0j} = \frac{\hat{\nu}_j \hat{\beta}_j \Lambda}{\mu(\hat{\nu}_j + \mu)(\hat{\gamma}_j + \mu)}$ , with  $\hat{\beta}_j$ ,  $\hat{\nu}_j$  and  $\hat{\gamma}_j$  given by 27, 26 and 23, respectively.  $\hat{R}_{0j}$  is graphed with respect to the time for smoothed COVID-19 data from Bogota.

<https://doi.org/10.1371/journal.pone.0285624.g021>



**Table 5.** Mean and standard deviation of update data estimations of the basic reproduction number,  $\hat{R}_{0j} = \frac{\hat{\nu}_j \hat{\beta}_j \Lambda}{\mu(\hat{\nu}_j + \mu)(\hat{\gamma}_j + \mu)}$  (with  $\hat{\beta}_j$ ,  $\hat{\nu}_j$  and  $\hat{\gamma}_j$  given by 27, 26 and 23, respectively) for smoothed COVID-19 data from Bogota in the eight intervals considered in the Fig 6.

Period ( $\tau_s$ )	Mean of $\hat{R}_{0j}$	SD of $\hat{R}_{0j}$	Mean of smoothed $\hat{R}_{0j}$	SD of smoothed $\hat{R}_{0j}$
[1,150)	18.353	150.219	19.318	22.827
[150,177)	1.267	23.723	-2.726	0.299
[177,205)	5.834	3.851	-0.942	0.764
[205,270)	5.122	13.458	3.879	2.014
[270,312)	4.133	1.591	8.249	0.587
[312,330)	1.417	13.168	9.019	0.017
[330,363)	5.848	10.689	8.806	0.097
[363,385)	16.710	13.879	9.761	0.904

<https://doi.org/10.1371/journal.pone.0285624.t005>

For interpreting the basic reproduction number, we can establish that the COVID-19 has invaded the population, and produced a state of endemicity in Bogota. If the infection conditions continue as in the period [363, 385), the epidemic could invade the population; unless control and mitigation measures like vaccination and isolation are applied. Actually, in Bogota, 55.2% of the population has been vaccinated according to Secretaría de Salud de Bogota data [46], what it would hope that  $R_0 < 1$ , if it is calculated for the data from august 24<sup>th</sup>/2021 to October 13<sup>rd</sup>/2021. On the other hand, in Table 5 can be seen that there is significant variability of  $\hat{R}_0$  in particular for the first period, which means that the infection conditions can change considerably between one day and another.

### SEIR model with random perturbations and its estimation

The above SIR and SEIR models are *deterministic*. However, epidemics tend to occur in cycles of outbreaks due to variations in the infection rate mainly related to certain external factors, such as people’s social activities and climatic fluctuations. The *climatic variations* can affect the *infection rate*  $\beta$ . According to [47], “many pathogens causing needle diseases are sensitive to precipitation and humidity, and their rates of reproduction, spread, and infection are greater when conditions are moist”. More recently reported, including media about the evidence of the mechanism by which climate change could have played a direct role in the emergence of SARs-CoV-2 [48–52]. Actually, [53] suggests that the climatology parameters could potentially affect the spread of the COVID-19. On the other hand, in [16] is highlight that the deterministic models “do not involve the variability of the sources of the information nor the possible errors and biases”, therefore, we model also the infection rate randomly. Some studies as [54–56] use the Brownian motion to model spatial-temporally as the temperature and weather variations affect the pollen dynamic, and the infection rate on epidemic; using equations similarly to 29 to some model parameters. This was applied even on partial differential systems. We consider the infection rate as a stochastic parameter (through random perturbations) and equation is given by

$$\tilde{\beta} := \beta + \sigma B(t), \tag{29}$$

Where  $\beta$  and  $\sigma$  are positive constants, and  $\{B(t)\}_{t \geq 0}$  is the *standard Brownian motion* on the probability space  $(\Omega, \mathfrak{F}, \mathbb{P},)$  which is driving the fluctuations in the dynamics of the epidemic. As we know that  $dB(t)$  is the increment of the standard Brownian motion and is normally distributed. Also, the parameter  $\beta$  is the rate of transmission of infection, and  $\sigma$  is the *volatility*

parameter which describes the amount of uncertainty of the parameter  $\beta$ . Now, we replace  $\beta dt$  by  $\beta dt + \sigma dB(t)$  in the system 1, we now propose the following system of stochastic differential equations for the SEIR model with random perturbations, whose system is the Eq 30.

$$\begin{aligned}
 dS(t) &= (\Lambda - \beta I(t)S(t) - \mu S(t))dt - \sigma I(t)S(t)dB(t) \\
 dE(t) &= (\beta I(t)S(t) - \nu E(t) - \mu E(t))dt + \sigma I(t)S(t)dB(t) \\
 dI(t) &= (\nu E(t) - \gamma I(t) - \mu I(t))dt \\
 dR(t) &= (\gamma I(t) - \mu R(t))dt,
 \end{aligned}
 \tag{30}$$

Our model considers environmental variations and social behaviors in the infection rate, inside of a Brownian motion with a volatility parameter. In [57] such variations are modeled by the Eq 29. One weakness of modeling infection rate by using random perturbations is that to big values of volatility parameter, it will have big values for  $E(t)$  or  $I(t)$ , therefore we expect small values when we implement it to real data. In the next subsection 1, an estimator of  $\sigma$  by minimizing the sum of ordinary squares and using the estimators of the susceptible and exposed population given by 11.

### Estimation of volatility parameter

We can approximate the populations under the SEIR model with random perturbations by using the approximations given by the Euler-Maruyama method [36]:

$$\begin{aligned}
 \tilde{S}(t_j) &= S(t_{j-1}) + [\Lambda - \beta S(t_{j-1})I(t_{j-1}) - \mu S(t_{j-1})](t_j - t_{j-1}) - \sigma S(t_{j-1})I(t_{j-1}) \\
 &\quad \cdot (B(t_j) - B(t_{j-1})) = \hat{S}(t_j) - \sigma S(t_{j-1})I(t_{j-1})(B(t_j) - B(t_{j-1})) \\
 \tilde{E}(t_j) &= E(t_{j-1}) + [\beta S(t_{j-1})I(t_{j-1}) - (\nu + \mu)E(t_{j-1})](t_j - t_{j-1}) + \sigma S(t_{j-1})I(t_{j-1}) \\
 &\quad \cdot (B(t_j) - B(t_{j-1})) = \hat{E}(t_j) + \sigma S(t_{j-1})I(t_{j-1})(B(t_j) - B(t_{j-1})) \\
 \tilde{I}(t_j) &= I(t_{j-1}) + [\nu E(t_{j-1}) - (\mu + \gamma)I(t_{j-1})](t_j - t_{j-1}) = \hat{I}(t_j) \\
 \tilde{R}(t_j) &= R(t_{j-1}) + [\gamma I(t_{j-1}) - \mu R(t_{j-1})](t_j - t_{j-1}) = \hat{R}(t_j),
 \end{aligned}
 \tag{31}$$

where  $\hat{S}(t_j), \hat{E}(t_j), \hat{I}(t_j)$  and  $\hat{R}(t_j)$  are the solutions of the deterministic system 30, in this case, approximated by Euler Method. We want to find the value of  $\sigma$  such as  $X$  is minimum, where  $X$  is

$$X = \sum_{j=0}^n \left( (\tilde{S}(t_j) - S(t_j))^2 + (\tilde{E}(t_j) - E(t_j))^2 + (\tilde{I}(t_j) - I(t_j))^2 + (\tilde{R}(t_j) - R(t_j))^2 \right). \tag{32}$$

As we do not have the data  $S(t_j)$  and  $E(t_j)$ , we take  $S(t_j) \approx \tilde{S}(t_j)$  and  $E(t_j) \approx \tilde{E}(t_j)$ , values calculated using  $\Lambda, \nu, \mu, I(t_j)$  and  $R(t_j)$ . For finding the minimum  $\sigma$ , we find  $\partial X / \partial \sigma$  given by

$$\begin{aligned}
 \frac{\partial X}{\partial \sigma} &= -2 \sum_{j=1}^{n+1} (\hat{S}(t_j) - \sigma \tilde{S}(t_{j-1})I(t_{j-1})\Delta_{B_j} - \tilde{S}(t_j))\tilde{S}(t_{j-1})I(t_{j-1})\Delta_{B_j} \\
 &\quad + 2 \sum_{j=1}^{n+1} (\hat{E}(t_j) + \sigma \tilde{S}(t_{j-1})I(t_{j-1})\Delta_{B_j} - \tilde{E}(t_j))\tilde{S}(t_{j-1})I(t_{j-1})\Delta_{B_j} = 0,
 \end{aligned}
 \tag{33}$$

taking  $\Delta t = t_j - t_{j-1}$  and  $\Delta_{B_j} = B(t_j) - B(t_{j-1})$  for all  $j = 1, \dots, n$ . We have that

$$2\sigma \sum_{j=1}^{n+1} \check{S}^2(t_{j-1})I^2(t_{j-1})\Delta_{B_j}^2 = \sum_{j=1}^{n+1} (\hat{S}(t_j) - \check{S}(t_j))\check{S}(t_{j-1})I(t_{j-1})\Delta_{B_j} + \sum_{j=1}^{n+1} (\check{E}(t_j) - \hat{E}(t_j))\check{S}(t_{j-1})I(t_{j-1})\Delta_{B_j}, \tag{34}$$

for  $\Delta_B \neq 0$ . Therefore, the estimator of  $\sigma$  is equal to

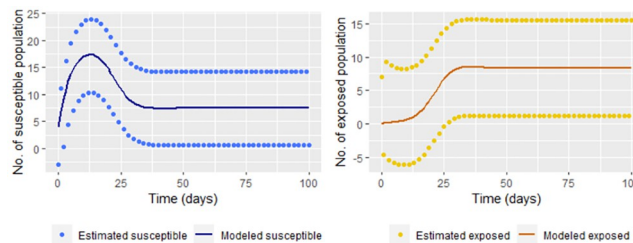
$$\hat{\sigma} = \frac{\sum_{j=1}^{n+1} (\hat{S}(t_j) - \check{S}(t_j) + \check{E}(t_j) - \hat{E}(t_j))\check{S}(t_{j-1})I(t_{j-1})\Delta_{B_j}}{2 \sum_{j=1}^{n+1} \check{S}^2(t_{j-1})I^2(t_{j-1})\Delta_{B_j}^2}, \tag{35}$$

where  $\Delta_{B_j}$  can be a random number generated of a distribution  $N(0, \Delta t)$ . We have that  $\partial^2 X / \partial \sigma^2 = 4 \sum_{j=1}^{n+1} \check{S}^2(t_{j-1})I^2(t_{j-1})\Delta_{B_j}^2 > 0$ , therefore  $\hat{\sigma}$  does  $X$  minimum. However, the values of  $\Delta_{B_j}$  values can significantly change which is why we will take  $\Delta_{B_j} \approx \bar{x}$  where  $\bar{x}$  is the mean of 10,000 random values of a distribution  $N(0, \Delta t)$ .

Worth noting that initially, we do not know the values of  $\beta$  and  $\gamma$ . Therefore these are estimated using the Eqs 15 and 17, respectively. There by  $\hat{S}(t_j)$ ,  $\hat{E}(t_j)$ ,  $\hat{I}(t_j)$  and  $\hat{R}(t_j)$  are the solutions of the system 1 with  $\beta \approx \hat{\beta}$  and  $\gamma \approx \hat{\gamma}$ , which we establish using the function `ode` from `deSolve` package in R.

Before estimating the parameters, we suggest smoothing the values of  $I(t_j)$  and  $R(t_j)$  (comparing Figs 11 and 12), which we do for the simulated data in the steps 1 and 2. With it, we have reasonable values of  $\check{S}(t_j)$  and  $\check{E}(t_j)$  (Eq 11) to calculate  $\hat{\sigma}$ . We use from day 7 to the last day for estimating  $\sigma$  due to  $\check{E}(t_j) \approx E(t_j)$  (Fig 10). In the Fig 22 we graph five paths for  $\check{S}(t_j)$ ,  $\check{E}(t_j)$ ,  $\hat{I}(t_j)$  and  $\hat{R}(t_j)$  (Eq 31) join the data simulated in the steps 1 and 2,  $\check{S}(t_j)$  and  $\check{E}(t_j)$  calculated of the smoothed simulated data. We obtained  $\hat{\sigma} = 0.07527228$ , which implies a great variability between the paths as can be observed in Fig 22.

On the Eq 31 the values of  $\hat{S}(t_j)$ ,  $\hat{E}(t_j)$ ,  $\hat{I}(t_j)$  and  $\hat{R}(t_j)$  are the solutions of the deterministic system 1 for COVID-19 data from Bogota. Given that we had a bad approximation by



**Fig 22. Five paths of the system 31 (Stochastic paths) with  $\Lambda = 4, v = 0.1, \mu = 0.2, \hat{\beta}, \hat{\gamma}$  and  $\hat{\sigma}$  (using 15, 17 and 35) calculated for the smoothed simulated data in the steps 1 and 2,  $\check{S}(t_j)$  and  $\check{E}(t_j)$ .** In this figure also are graphed the solution of the deterministic system 1 (Deterministic solution) with  $\Lambda, v$  and  $\mu, \beta,$  and  $\gamma$  previously given and the simulated data in the steps 1 and 2 (Infected and Recovered data),  $\check{S}(t_j)$  (Estimated susceptible) and  $\check{E}(t_j)$  (Estimated exposed).  $\check{S}(t_j)$  and  $\check{E}(t_j)$  are calculated for the smoothed simulated data.

<https://doi.org/10.1371/journal.pone.0285624.g022>

estimating using least squares (Table 3), we decided to take  $\hat{S}(t_j), \hat{E}(t_j), \hat{I}(t_j)$  and  $\hat{R}(t_j)$  by

$$\underline{S}(t_j) = N(t_j) - \widehat{E}(t_j) - \widehat{I}(t_j) - \widehat{R}(t_j), \underline{E}(t_j) = \widehat{E}(t_j), \underline{I}(t_j) = \widehat{I}(t_j) \text{ and } \underline{R}(t_j) = \widehat{R}(t_j); \quad (36)$$

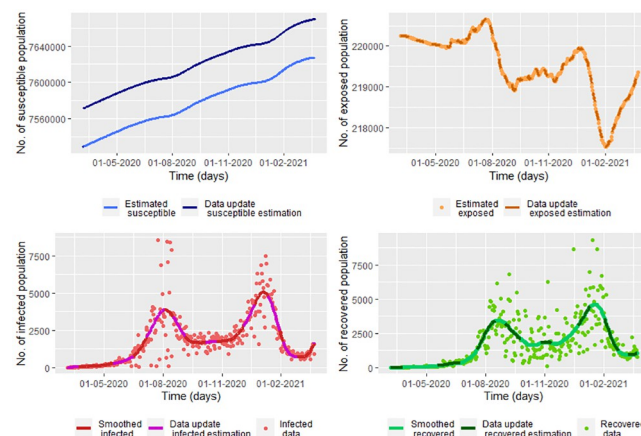
These are the updated data estimations graphed in Fig 18, except for  $\underline{S}(t_j)$  whose estimations are not good using this approach (see Fig 18). It is important to say that  $\underline{S}(t_j), \underline{E}(t_j), \underline{I}(t_j)$  and  $\underline{R}(t_j)$  are calculated from  $\nu = 1/5.2, \mu = 4/1000, \Lambda = 73660$  and smoothed infected and recovered COVID-19 data (which we note as  $I(t_j)$  and  $R(t_j)$ , respectively). Thus we estimate the volatility parameter as

$$\hat{\sigma} = \frac{\sum_{j=1}^{n+1} (\underline{S}(t_j) - \check{S}(t_j) + \check{E}(t_j) - \underline{E}(t_j)) \check{S}(t_{j-1}) I(t_{j-1}) \Delta_{B_j}}{2 \sum_{j=1}^{n+1} \check{S}^2(t_{j-1}) I^2(t_{j-1}) \Delta_{B_j}^2} = 6.756712 \times 10^{-23}$$

Which is negligible by being too near to zero. This is why  $\underline{S}(t_j), \underline{E}(t_j)$  are practically equal to  $\check{S}(t_j), \check{E}(t_j)$ , respectively (see Figs 18 and 19). For this reason, we take  $\hat{S}(t_j), \hat{E}(t_j), \hat{I}(t_j)$  and  $\hat{R}(t_j)$ , respectively by  $\widehat{S}(t_j), \widehat{E}(t_j), \widehat{I}(t_j)$  and  $\widehat{R}(t_j)$  defined as in the Eq 28. We have  $\hat{\sigma}$  given by

$$\hat{\sigma} = \frac{\sum_{j=1}^{n+1} (\widehat{S}(t_j) - \check{S}(t_j) + \check{E}(t_j) - \widehat{E}(t_j)) \check{S}(t_{j-1}) I(t_{j-1}) \Delta_{B_j}}{2 \sum_{j=1}^{n+1} \check{S}^2(t_{j-1}) I^2(t_{j-1}) \Delta_{B_j}^2} = 7.193334 \times 10^{-7} \quad (37)$$

In Fig 23, we graph the solutions for the susceptible and exposed population given by  $\underline{S}(t_j)$  and  $\widehat{E}(t_j)$  defined as in the Eq 28, join with the stochastic paths under the stochastic model 4 (system 38). Note that we take  $\underline{S}(t_j)$  instead of  $\widehat{S}(t_j)$  due to with  $\widehat{S}(t_j)$  there is not a good fit of  $\check{S}(t_j)$  (upper left Fig 18). In the Fig 23 the maximum error of prediction, taking the distance between the estimated susceptible population given by  $\check{S}(t_j)$  and the paths of the stochastic system 38, for the susceptible population is approximately 77,794, and the maximum prediction



**Fig 23. Five paths of the system 38 for susceptible and exposed population.** We take  $\Lambda = 73660, \mu = 0.004, \hat{\beta}_j, \hat{\nu}_j$  and  $\hat{\gamma}_j$  and  $\hat{\sigma}$  (using 27, 26, 23 and 37) calculated for the smoothed COVID-19 data from Bogota,  $\check{S}(t_j)$  and  $\check{E}(t_j)$ .  $\check{S}(t_j)$  and  $\check{E}(t_j)$  are calculated for the smoothed data.

<https://doi.org/10.1371/journal.pone.0285624.g023>

error for the exposed population is approximately 60,153.

**Model 4 :**

$$\begin{aligned} \tilde{S}(t_j) &= \underline{S}(t_j) - \hat{\sigma} \check{S}(t_{j-1}) I(t_{j-1}) (B(t_j) - B(t_{j-1})) \\ \tilde{E}(t_j) &= \widehat{E}(t_j) + \hat{\sigma} \check{S}(t_{j-1}) I(t_{j-1}) (B(t_j) - B(t_{j-1})) \\ \tilde{I}(t_j) &= \widehat{I}(t_j), \tilde{R}(t_j) = \widehat{R}(t_j). \end{aligned} \tag{38}$$

In the equation, 38, we have a deterministic graph for the infected and recovered population. So, there is no error prediction. In this way, we propose taking  $\tilde{I}(t_j) = N(t_j) - \tilde{S}(t_j) - \tilde{E}(t_j) - \tilde{R}(t_j)$  and  $\tilde{R}(t_j) = N(t_j) - \tilde{S}(t_j) - \tilde{E}(t_j) - \tilde{I}(t_j)$  for having that error, whose paths graph in the Fig 24. Maximum error of prediction for the infected population for the infected population is approximately 502, and the maximum prediction error for the recovered population is the same. In this case, the distance between the smoothed infected COVID-19 data from Bogota and the paths of  $N(t_j) - \tilde{S}(t_j) - \tilde{E}(t_j) - \tilde{R}(t_j)$  with  $\tilde{S}(t_j) - \tilde{E}(t_j)$  from the system 38

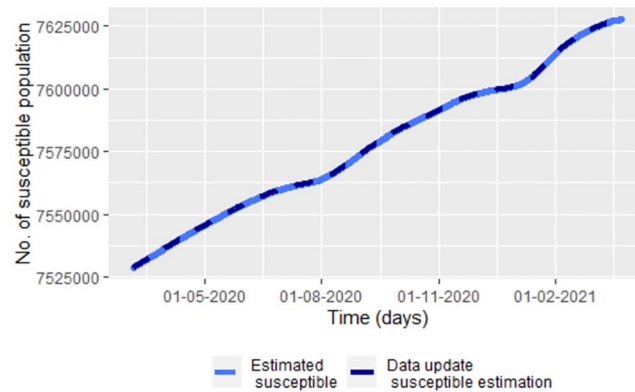
We study the homoscedasticity according to the graphs of fitted values of the populations concerning the residuals, which correspond to the Fig 25. We do a residual analysis under the regression focus using the model given by the system 39. In Fig 25, we have heteroscedasticity for all the populations under the regression model provided by the system 39, where it is assumed that the residuals for each population are independent.

$$\begin{aligned} \check{S}(t_j) &= \widehat{S}(t_j) + \epsilon_s, \epsilon_s \sim N(0, \sigma_s^2) \\ \check{E}(t_j) &= \widehat{E}(t_j) + \epsilon_e, \epsilon_e \sim N(0, \sigma_e^2) \\ I(t_j) &= \tilde{I}(t_j) + \epsilon_i, \epsilon_i \sim N(0, \sigma_i^2) \\ R(t_j) &= \tilde{R}(t_j) + \epsilon_r, \epsilon_r \sim N(0, \sigma_r^2) \end{aligned} \tag{39}$$

Table 6 shows the p-values obtained for the Shapiro-Wilks test for normality for all the residuals given the paths under the model 39. Note that all the p-values are more significant than 0.05, so we conclude with a confidence of 95% that the residuals do not present normal distribution.

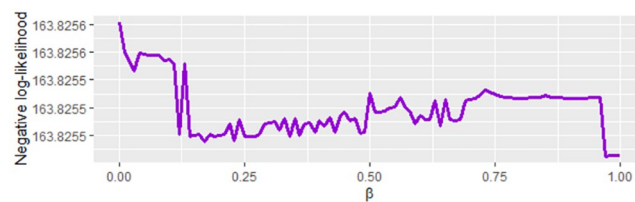
**Conclusions**

This paper proposes estimation methods using a data update approach for the COVID-19 data in Bogota. Our methods of estimation based on recovered as infected data: (1) a method based on the likelihood function with variance given by the Eq 6 (model 2); (2) a method based on ordinary least squares on the infected and recovered data given by the Eq 7; and (3) a data update method based on the recovered, infected and exposed data; given by the Eqs 23, 26 and 27 (model 3). Method (1) has the issue that other populations can be overestimated. In contrast, the other methods base their estimation on first approximating the susceptible and exposed population using some known parameters, which are  $\Lambda, \mu,$  and  $v$ . In particular, method (1) could be convenient for the initial phase of increasing the disease; and when there



**Fig 24.** Five paths of  $\tilde{I}(t_j) = N(t_j) - \tilde{S}(t_j) - \tilde{E}(t_j) - \tilde{R}(t_j)$  and  $\tilde{R}(t_j) = N(t_j) - \tilde{S}(t_j) - \tilde{E}(t_j) - \tilde{I}(t_j)$  with  $\tilde{S}(t_j)$  and  $\tilde{E}(t_j)$  from the system 38 for infected and recovered population. We take  $\Lambda = 73660, \mu = 0.004, \hat{\beta}_j, \hat{\nu}_j$  and  $\hat{\gamma}_j$  and  $\hat{\sigma}$  (using 27, 26, 23 and 37) calculated for the smoothed COVID-19 data from Bogota,  $\check{S}(t_j)$  and  $\check{E}(t_j)$ .  $S(t_j)$  and  $E(t_j)$  are calculated for the smoothed data.

<https://doi.org/10.1371/journal.pone.0285624.g024>



**Fig 25.** Fitted values for each population concerning the residuals according to the model 39.

<https://doi.org/10.1371/journal.pone.0285624.g025>

is not enough knowledge about, for instance, how fast the epidemic spreads. However, their estimators may change significantly by small changes in the initial conditions. Method (2) does not always fit the data, and depending on the known parameters, this method may present a different scale in their solutions. It is worth noting that although model (2) has a good fit, we consider that the model methodology may not determine the exceeded cases.

The proposed models capture the data from Bogotá city well. However, it is potentially limited by the lack of prevalence data corresponding to the registration in individuals who present specific variants of the disease over a period of time, including susceptible, exposed, and vaccinated sub-populations, and also data of asymptomatic infected. Also, we have not considered age-structured models in social behaviors and transmission rates. We create the data update approach for a good fit. Under the update data method, we can think of the parameters as functions depending on the time or a time-series model of the ARIMA process. However, the

**Table 6.** p-values for each one of the paths under the model 39 under the Shapiro-Wilks test for normality.

Population	c p-value for the 1st path	c p-value for the 2nd path	c p-value for the 3rd path	c p-value for the 4th path	c p-value for the 5th path
Susceptible	1.892e-13	6.565e-15	1.624e-14	6.224e-18	1.671e-14
Exposed	1.789e-13	6.259e-15	1.631e-14	5.988e-18	1.739e-14
Infected	2.207e-24	2.207e-24	2.207e-24	2.207e-24	2.207e-24
Recovered	6.344e-18	6.344e-18	6.344e-18	6.344e-18	6.344e-18

<https://doi.org/10.1371/journal.pone.0285624.t006>



parameter  $v$  must be newly estimated under this. We concluded that the best method for fitting an epidemic mathematical model to infected and recovered COVID-19 data from Bogota, D. C. is by using the model 3 and taking the susceptible population as  $\underline{S}(t_j)$ . If we wished to make trusted bands, we could estimate the parameters by the Eqs 23, 26 and 27 and approximate by a time-series process for predicting the behavior of an epidemic based on the infected and recovered data. This paper has established some methodologies of parameter estimation on models based on ordinary differential systems and stochastic differential systems to one of the simplest models: the SEIR model. In future research, we expect to develop the data update approach to models with more compartments, including symptomatic, hospitalized, and dead for the COVID population.

## Supporting information

**S1 File.**  
(PDF)

## Acknowledgments

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

## Author Contributions

**Conceptualization:** Viswanathan Arunachalam.

**Investigation:** Andrés Ríos-Gutiérrez, Soledad Torres, Viswanathan Arunachalam.

**Methodology:** Andrés Ríos-Gutiérrez, Viswanathan Arunachalam.

**Validation:** Andrés Ríos-Gutiérrez, Soledad Torres.

**Writing – original draft:** Andrés Ríos-Gutiérrez, Viswanathan Arunachalam.

**Writing – review & editing:** Soledad Torres.

## References

1. World Health Organization. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). [https://www.who.int/es/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/es/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)), 2020.
2. Semana. Coronavirus en Colombia: primer caso confirmado. <https://www.semana.com/nacion/articulo/coronavirus-en-colombia-primer-caso-confirmado/655252>, 2020.
3. Instituto Nacional de Salud. COVID-19 en Colombia. <https://www.ins.gov.co/Noticias/paginas/coronavirus.aspx> & 2023.
4. Alcaldía Mayor de Bogotá. Tres nuevas medidas para combatir el COVID-19 en Bogotá. <https://bogota.gov.co/mi-ciudad/salud/cuarentena/tres-nuevas-medidas-para-combatir-el-covid-19-en-bogota>, 2020.
5. Presidencia de la República de Colombia. Decretos acerca del COVID-19 en Colombia. <https://coronaviruscolombia.gov.co/Covid19/decretos.html>, 2020 & 2021.
6. Ministerio de Salud y de la Protección Social. Plan Nacional de Vacunación contra el COVID-19. Información oficial la vacunación contra el coronavirus en Colombia. <https://www.minsalud.gov.co/salud/publica/Vacunacion/Paginas/Vacunacion-covid-19.aspx>, 2023.
7. Kermack WO, McKendrick AG. Contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115 (772):700–721, 1927.

8. Centers for Disease Control and Prevention. Coronavirus disease 2019 (COVID-19). <https://www.cdc.gov/coronavirus/2019-ncov/index.html>, 2020.
9. Korobeinikov A. Global properties of SIR and SEIR epidemic models with multiple parallel infectious stages. *Bulletin of mathematical biology*, 71(1):75–83, 2009. <https://doi.org/10.1007/s11538-008-9352-z> PMID: 18769976
10. Liu W, Mao X. Strong convergence of the stopped Euler-Maruyama method for nonlinear stochastic differential equations. *Applied Mathematics and Computation*, 223:389–400, 2013. <https://doi.org/10.1016/j.amc.2013.08.023>
11. López L, Rodo X. A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics. *Results in Physics*, 21:103746, 2021. <https://doi.org/10.1016/j.rinp.2020.103746> PMID: 33391984
12. Mwalili S, Kimathi M, Ojiambo V, Gathungu D, Mbogo R. SEIR model for COVID-19 dynamics incorporating the environment and social distancing. *BMC Research Notes*, 13(1):1–5, 2020. <https://doi.org/10.1186/s13104-020-05192-1> PMID: 32703315
13. Ryad G, Boone E, Abdel-Salam AS. SEIRD model for Qatar COVID-19 outbreak: a case study *Letters in Biomathematics*, 8(1):19–28, 2021.
14. Suwardi A, Muh Isbar P, Muh R, Wahidah S, Syafruddin S. Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia. *Chaos, Solitons & Fractals*, 139:110072, 2020. <https://doi.org/10.1016/j.chaos.2020.110072>
15. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of thoracic disease*, 12(3):165, 2020. <https://doi.org/10.21037/jtd.2020.02.64> PMID: 32274081
16. Cuesta-Herrera L, Pastenes L, Córdova-Lepe F, Arencibia AD, Torres-Mantilla H, Gutiérrez-Jara JP. Analysis of seir-type models used at the beginning of covid-19 pandemic reported in high-impact journals, 2022.
17. Gomez J, Prieto J, Leon E, Rodríguez A. Infekta—an agent-based model for transmission of infectious diseases: The covid-19 case in Bogotá, Colombia. *PloS one*, 16(2):e0245787, 2021. <https://doi.org/10.1371/journal.pone.0245787> PMID: 33606714
18. Niño-Torres D, Ríos-Gutiérrez A, Arunachalam V, Ohajunwa C, Seshaiyer P. Stochastic modeling, analysis, and simulation of the COVID-19 pandemic with explicit behavioral changes in Bogotá: A case study. *Infectious Disease Modelling*, 7(1):199–211, 2022. <https://doi.org/10.1016/j.idm.2021.12.008> PMID: 35005324
19. Rivera-Rodríguez C, Urdinola BP. Predicting hospital demand during the COVID-19 outbreak in bogotá, colombia. *Frontiers in public health*, 8:582706, 2020. <https://doi.org/10.3389/fpubh.2020.582706> PMID: 33262969
20. Aliou MA, Baldé T. Fitting SIR model to COVID-19 pandemic data and comparative forecasting with machine learning. *BMJ*, 2020.
21. Chae SY, Lee KE, Lee HM, Jung N, Le QA, Mafwele BJ, et al. Estimation of infection rate and predictions of disease spreading based on initial individuals infected with COVID-19. *Frontiers in Physics*, page 311, 2020. <https://doi.org/10.3389/fphy.2020.00311>
22. Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, 139:110057, 2020. <https://doi.org/10.1016/j.chaos.2020.110057> PMID: 32834610
23. Vasconcelos GL, Cordeiro LP, Duarte-Filho GC, Brum AA. Modeling the epidemic growth of preprints on COVID-19 and SARS-CoV-2. *Frontiers in Physics*, 9:125, 2021. <https://doi.org/10.3389/fphy.2021.603502>
24. Yang HM, Lombardi-Junior LP, Yang AC. Are the SIR and SEIR models suitable to estimate the basic reproduction number for the COVID-19 epidemic? *medRxiv*, 2020.
25. Wu K, Darcet D, Wang Q, Sornette D. Generalized logistic growth modeling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world. *Nonlinear dynamics*, 101(3):1561–1581, 2020. <https://doi.org/10.1007/s11071-020-05862-6> PMID: 32836822
26. Zhan C, Situ W, Yeung LF, Tsang PWM, Yang G. A parameter estimation method for biological systems modelled by ODE/DDE models using spline approximation and differential evolution algorithm. *IEEE/ACM transactions on computational biology and bioinformatics*, 11(6):1066–1076, 2014. <https://doi.org/10.1109/TCBB.2014.2322360> PMID: 26357044
27. Faranda D, Alberti T. Modeling the second wave of COVID-19 infections in France and Italy via a stochastic SEIR model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(11):111101, 2020. <https://doi.org/10.1063/5.0015943>
28. Hosseini PR, Dhondt AA, Dobson A. Seasonality and wildlife disease: how seasonal birth, aggregation and variation in immunity affect the dynamics of *Mycoplasma gallisepticum* in house finches.

- Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1557):2569–2577, 2004. <https://doi.org/10.1098/rspb.2004.2938> PMID: 15615682
29. Martelloni G, Martelloni G. Modelling the downhill of the SARS-CoV-2 in Italy and a universal forecast of the epidemic in the world. *Chaos, Solitons & Fractals*, 139:110064, 2020. <https://doi.org/10.1016/j.chaos.2020.110064>
  30. Tsay C, Lejarza F, Stadtherr MA, Baldea M. Modeling, state estimation, and optimal control for the US COVID-19 outbreak. *Scientific Reports*, 10(1):1–12, 2020. <https://doi.org/10.1038/s41598-020-67459-8> PMID: 32612204
  31. Marsili-Libelli S, Guerrizio S, and Checchi N. Confidence regions of estimated parameters for ecological systems. *Ecological Modelling*, 165(2-3):127–146, 2003. [https://doi.org/10.1016/S0304-3800\(03\)00068-1](https://doi.org/10.1016/S0304-3800(03)00068-1)
  32. Butler D. Models overestimate Ebola cases. *Nature News*, 515(7525):18, 2014. <https://doi.org/10.1038/506018a> PMID: 25373654
  33. Platt DE, Parida L, Zalloua P. Lies, gosh darn lies, and not enough good statistics: why epidemic model parameter estimation fails. *Scientific Reports*, 11(1):1–10, 2021. <https://doi.org/10.1038/s41598-020-79745-6>
  34. Departamento Administrativo Nacional de Estadística (DANE). Documento metodológico de elaboración de las proyecciones de población de Bogotá, D.C., a nivel de localidad hasta el año 2035 y de Unidad de Planeamiento Zonal—UPZ hasta el año 2024. <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion/proyecciones-de-poblacion-bogota>, 2020.
  35. King AA, Ionides EL, Pascual M, Bouma MJ. Inapparent infections and cholera dynamics. *Nature*, 454(7206):877–880, 2008. <https://doi.org/10.1038/nature07084> PMID: 18704085
  36. Iacus SM. *Simulation and inference for stochastic differential equations: with R examples*, volume 486. Springer, 2008.
  37. Cheng C, Zhang DD, Dang D, Geng J, Zhu P, Yuan M, et al. The incubation period of COVID-19: a global meta-analysis of 53 studies and a Chinese observation study of 11545 patients. *Infectious Diseases of Poverty*, 10(1):1–13, 2021. <https://doi.org/10.1186/s40249-021-00901-9>
  38. Ministerio de Relaciones Exteriores. Estadísticas de Migración Colombia. <https://www.migracioncolombia.gov.co/planeacion/estadisticas>, 2023.
  39. Maleki M, Mahmoudi MR, Wraith D, Pho KH. Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel Medicine and Infectious Disease*, 37:101742, 2020. <https://doi.org/10.1016/j.tmaid.2020.101742>
  40. Gianfranco A, Giuseppe R, Stefano C, Alberto G, La Vecchia C. Excess total mortality during the COVID-19 pandemic in Italy: updated estimates indicate persistent excess in recent months. *La Medicina del Lavoro*, 113(2), 2022.
  41. Irish A, Whitman JD, Clark EH, Marcus R, Bern C. Updated estimates and mapping for prevalence of Chagas disease among adults, United States. *Emerging Infectious Diseases*, 28(7):1313, 2022. <https://doi.org/10.3201/eid2807.212221> PMID: 35731040
  42. Tang B, Bragazzi NL, Li Q, Tang S, Xiao Y, Wu J. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCoV). *Infectious Disease Modelling*, 5:248–255, 2020. <https://doi.org/10.1016/j.idm.2020.02.001> PMID: 32099934
  43. Beck JL, Katafygiotis LS. Updating models and their uncertainties. I: Bayesian statistical framework. *Journal of Engineering Mechanics*, 124(4):455–461, 1998. [https://doi.org/10.1061/\(ASCE\)0733-9399\(1998\)124:4\(455\)](https://doi.org/10.1061/(ASCE)0733-9399(1998)124:4(455))
  44. Ríos-Gutiérrez A, Torres S, Arunachalam V. Studies on the basic reproduction number in stochastic epidemic models with random perturbations. *Advances in Difference Equations*, 2021(1):1–24, 2021. <https://doi.org/10.1186/s13662-021-03445-2> PMID: 34149835
  45. Foppa IM. *A historical introduction to mathematical modeling of infectious diseases: Seminal Papers in Epidemiology*. Academic Press, 2016.
  46. Secretaría Distrital de Salud. Vacunación contra COVID-19 en Bogotá DC. <https://saludata.saludcapital.gov.co/osb/index.php/datos-de-salud/enfermedades-trasmisibles/covid-19-vacunometro>, 2022.
  47. Sturrock RN, Frankel SJ, Brown AV, Hennon PE, Kliejunas JT, Lewis KJ, et al. Climate change and forest diseases. *Plant Pathology*, 60(1):133–149, 2011. <https://doi.org/10.1111/j.1365-3059.2010.02406.x>
  48. Beyer RM, Manica A, Mora C. Shifts in global bat diversity suggest a possible role of climate change in the emergence of SARS-CoV-1 and SARS-CoV-2. *Science of The Total Environment*, page 145413, 2021. <https://doi.org/10.1016/j.scitotenv.2021.145413> PMID: 33558040

49. Md NS, Wai YC, Ibrahim N, Rashid ZZ, Mustafa N, Hamid HHA, et al. Particulate matter (PM2.5) as a potential SARS-CoV-2 carrier. 2020.
50. Tung NT, Cheng PC, Chi KH, Hsiao TC, Jones T, Kelly BÉruBÉ, et al. Particulate matter and SARS-CoV-2: a possible model of covid-19 transmission. *Science of The Total Environment*, 750:141532, 2021. <https://doi.org/10.1016/j.scitotenv.2020.141532> PMID: 32858292
51. Patz JA, Githeko AK, McCarty JP, Hussein S, Confalonieri U, Wet ND, et al. Climate change and infectious diseases. *Climate change and human health: risks and responses*, 6:103–137, 2003.
52. Rajagopalan S, Huang S, Brook RD. Flattening the curve in COVID-19 using personalised protective equipment: lessons from air pollution, 2020.
53. Ali B, Golafshani EM, Hosseini SM. Determinants of the infection rate of the COVID-19 in the us using anfis and virus optimization algorithm (voa). *Chaos, Solitons & Fractals*, 139:110051, 2020. <https://doi.org/10.1016/j.chaos.2020.110051>
54. Laredo C, Grimaud A. Stochastic models and statistical inference for plant pollen dispersal. *Journal de la société française de statistique*, 148(1), 2007.
55. Zhou Y, Zhang W, Yuan S. Survival and stationary distribution of a SIR epidemic model with stochastic perturbations. *Applied Mathematics and Computation*, 244:118–131, 2014. <https://doi.org/10.1016/j.amc.2014.06.100>
56. Ji C, Jiang D, Yang Q, Shi N. Dynamics of a multigroup SIR epidemic model with stochastic perturbation. *Automatica*, 48(1):121–131, 2012. <https://doi.org/10.1016/j.automatica.2011.09.044>
57. Gray A, Greenhalgh D, Hu L, Mao X, Pan J. A stochastic differential equation SIS epidemic model. *SIAM Journal on Applied Mathematics*, 71(3):876–902, 2011. <https://doi.org/10.1137/10081856X>