RESEARCH ARTICLE

On mobility trends analysis of COVID–19 dissemination in Mexico City

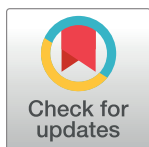**Kernel Prieto**[1☺]*, **M. Victoria Chávez–Hernández**[2☺], **Jhoana P. Romero–Leiton**[3☺]

1 Instituto de Matemáticas, Universidad Nacional Autónoma de México, Mexico, México, **2** Facultad de Ingeniería Mecánica y Eléctrica, Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Mexico, México, **3** Facultad de Ingeniería, Universidad Cesmag, Pasto, Colombia

☺ These authors contributed equally to this work.
* kernel@ciencias.unam.mx

Abstract

This work presents a tool for forecasting the spread of the new coronavirus in Mexico City, which is based on a mathematical model with a metapopulation structure that uses Bayesian statistics and is inspired by a data-driven approach. The daily mobility of people in Mexico City is mathematically represented by an origin-destination matrix using the open mobility data from Google and the Transportation Mexican Survey. This matrix is incorporated in a compartmental model. We calibrate the model against borough-level incidence data collected between 27 February 2020 and 27 October 2020, while using Bayesian inference to estimate critical epidemiological characteristics associated with the coronavirus spread. Given that working with metapopulation models leads to rather high computational time consumption, and parameter estimation of these models may lead to high memory RAM consumption, we do a clustering analysis that is based on mobility trends to work on these clusters of borough separately instead of taken all of the boroughs together at once. This clustering analysis can be implemented in smaller or larger scales in different parts of the world. In addition, this clustering analysis is divided into the phases that the government of Mexico City has set up to restrict individual movement in the city. We also calculate the reproductive number in Mexico City using the next generation operator method and the inferred model parameters obtaining that this threshold is in the interval (1.2713, 1.3054). Our analysis of mobility trends can be helpful when making public health decisions.

Introduction

The coronavirus disease 2019 (COVID-19) is caused by a novel coronavirus. The coronaviruses are a family of viruses that cause infection in humans and animals. The diseases that are by a coronavirus are zoonotic [1]. In particular, the coronaviruses that affect humans (HCoV) can produce clinical symptoms, such as the Severe Acute Respiratory Syndrome (SARS) viruses and Middle East Respiratory Syndrome (MERS-CoV) [2]. COVID-19 was first identified amid an outbreak of respiratory illness cases in Wuhan City, Hubei Province, China. This disease was initially reported to the WHO on 31 December 2019. On 11 March 2020, the

WHO declared COVID-19 to be a global pandemic [3]. From the beginning of the epidemic to 21 January 2021, more than 97,890,676 cases and 2,094,459 deaths have been reported globally.

The first case of COVID-19 in South America was registered in Brazil on 26 February 2020. The first death from this infection in this region was announced in Argentina on 7 March 2020. The virus then arrived in Mexico, where by 21 January 2021 there have been almost 1,688,944 confirmed cases and 144,371 deaths.

To date, many researchers around the world have focused on understanding the transmission dynamics of COVID-19 disease using mathematical and statistical models and methods, see for example [4–12]. In this work, we will focus on those models that incorporate information on human movement. The relationship between human mobility and the transmission of coronavirus disease in the United States has been studied in [13, 14]. Metapopulation models are not only among the simplest spatial models but they are also the most applicable to modelling many human diseases [15]. The metapopulation concept is to subdivide the entire population into distinct *sub-populations*, each of which has independent dynamics together with limited interaction between the sub-populations. This approach has been used to great effect within the ecological literature [16] and it has recently been used to model the spread of COVID-19; see, for example, [17–21].

In this work, we calibrate the metapopulation model proposed by Li et al. in [21], similar to [22], using incidence data reported in [23]. Consequently, we first describe the mathematical model that we have used; then, we compute the number of trips that are produced and attracted in each borough of Mexico City using data about these trips in 2017 [24], which then we combine with the rates of reduction or increase in mobility during the pandemic reported by Google [25] and the government of Mexico City [26]. Later, by using Bayesian inference, we solve the associated inverse problem to predict the dynamics of the spread of cases, similar to the following references [27–33]. Our conclusions are presented in the last section.

Computation of the mobility matrices

To incorporate mobility in the transmission model, the produced and attracted trips in the boroughs of Mexico City are considered (see Fig 1 and Table 1). Mexico City is the capital of Mexico. It has around 9 million inhabitants and a floating population of over 22 million, who are composed of daily commuters and international visitors. Mexico City is among the top 10 most crowded cities in the world [34]. It also has a large number of corporate headquarters and a large transport network, which is composed of 20 different modes of transport.

Mobility between the zones in Table 1 is represented in a two-dimensional arrangement, which is known as the origin-destination matrix (O-D matrix) $\mathbf{M} = \{M_{ij}\}$, $i, j = 1, \ldots, 16$, where M_{ij} represents the number of trips from zone i to zone j. Origin-destination matrices are usually obtained every 10 years from surveys. In Mexico City, the last survey was carried out in 2017 [24]. The information available identifies, among other things: if the trip was made on a weekday or if it was made during the weekend, the transport mode used, the purpose and the time. It is important to notice that, due to the complexity of mobility in Mexico City, the O-D matrix does not have to be symmetric nor the sum of the row i has to be equal to the sum of the column i. This can be explained because the O-D matrix captures chains of trips that may begin one day and end within the next few days. For instance, trips of people who leave their home to go to zone i to work and then go to zone j to see a movie before returning home, or people that work in zone i for few consecutive days (see [35]). In this paper, we consider all of the trips and they are identified by the mode of transport that was used in the area of interest (see Table 2).
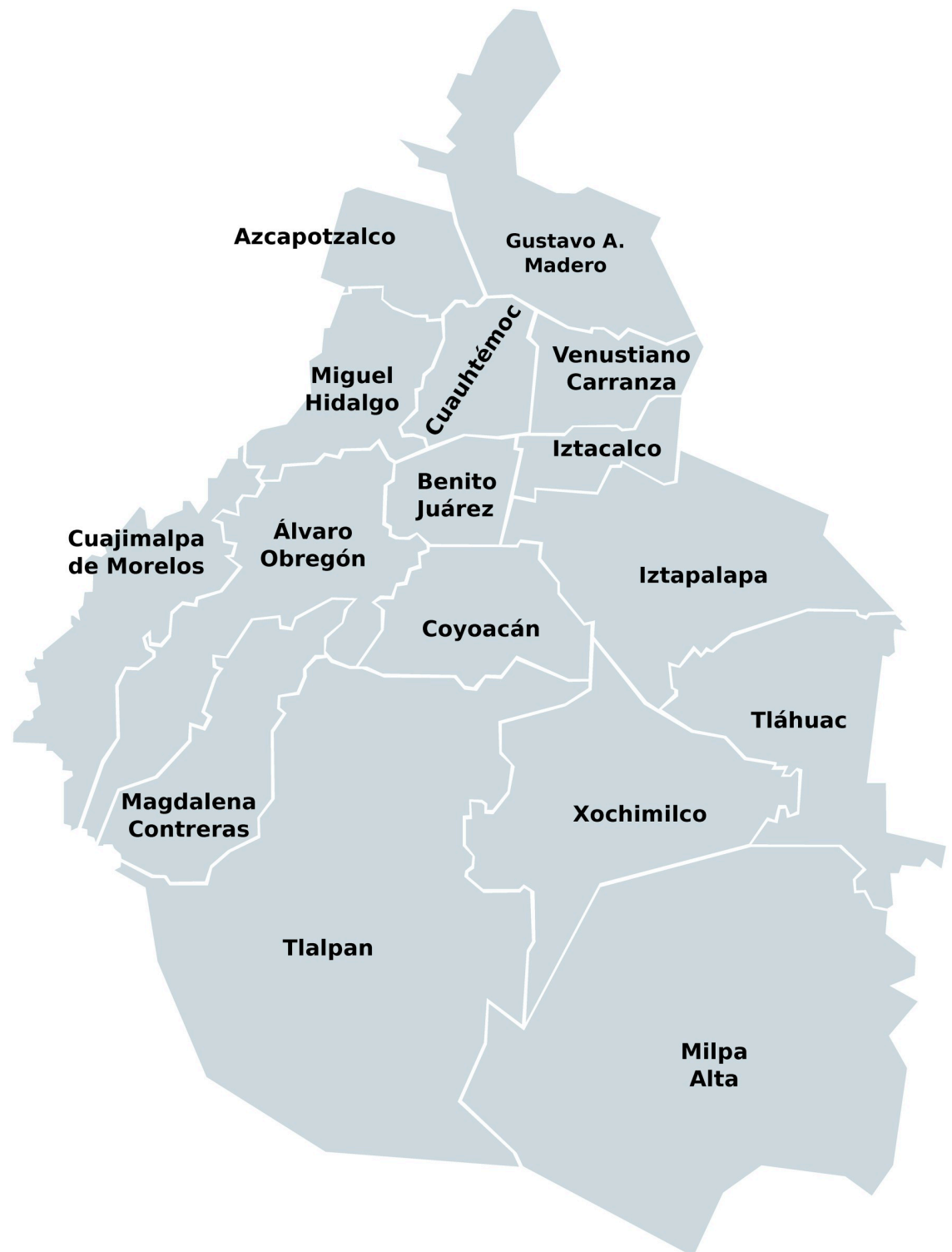
Fig 1. Boroughs of Mexico City.

Table 1. Boroughs of Mexico City.

Id.	Name	Id.	Name	Id.	Name	Id.	Name
1	Azcapotzalco	5	Iztacalco	9	Álvaro Obregón	14	Cuauhtémoc
2	Coyoacán	6	Iztapalapa	10	Tláhuac	15	Miguel Hidalgo
3	Cuajimalpa de Morelos	7	La Magdalena Contreras	11	Tlalpan	16	Venustiano Carranza
				12	Xochimilco		
4	Gustavo A. Madero	8	Milpa Alta	13	Benito Juárez		

https://doi.org/10.1371/journal.pone.0263367.t001

Table 2. Transport modes used in the area of interest.

Car	RTP/M1	Metrobus/Mexibus	Motorcycle taxi
Collective/micro	Bicycle	Light rail	School transportation
Taxi app	Bus	Suburban	Personal transportation
Taxi	Motorcycle	Walk	Other
Subway	Trolleybus	Pedicab	

https://doi.org/10.1371/journal.pone.0263367.t002

There are several methodologies to update the O-D matrix in the literature. Most of them combine known information with current data observed, such as the number of trips in some segments of the transit network [36]. There are also some approaches that project the trips to/from each zone based on the projected economic growth in those areas [37]. Nevertheless, given the pandemic situation that we are experiencing today and that we have current available data about the increase or decrease in mobility for some modes of transport, transit stations and parking lots, we consider the 2017 O-D matrix as a reference matrix and we update it to a scenario in 2020 using the daily mobility reports provided by Google [25] and the government of Mexico City [26]. According to [24], Tables 3 to 6 represent the number of trips between

Table 3. Mean number of trips from zones 1-16 to zones 1-9 during a week day.

Id	1	2	3	4	5	6	7	8	9
1	477560	21108	7564	120676	6475	25026	1557	0	24149
2	21111	886428	11010	48039	32230	228272	49305	17446	108741
3	7336	10462	301805	9770	5397	10867	3491	722	100180
4	120735	50796	9699	1677356	23427	52676	2602	590	27846
5	6486	32393	5358	22936	361938	181660	1563	145	19190
6	25244	221750	10308	56688	177845	2766471	9326	10395	64521
7	1685	50941	4148	2576	1389	8008	260789	441	83675
8	0	16022	722	590	145	10872	336	218175	3901
9	24180	108129	102892	28836	18012	61990	80171	5577	1034673
10	1730	50969	1994	7283	8243	134119	2179	22801	11070
11	11034	237367	10663	17958	14370	82757	53911	17203	95685
12	2148	100785	2637	6062	7586	46856	7030	31284	14993
13	17200	126058	14645	59944	59956	163732	29187	9714	161543
14	88252	118292	20292	264298	84725	253810	22415	9495	107543
15	99868	52246	43942	81013	26450	73629	11072	2972	115902
16	15795	31971	3324	103651	56192	102397	3239	1199	22583
Sum	920364	2115717	551003	2507676	884380	4203142	538173	348159	1996195

https://doi.org/10.1371/journal.pone.0263367.t003

Table 4. Mean number of trips from zones 1-16 to zones 10-16 during a week day.

Id	10	11	12	13	14	15	16	Sum
1	1918	10139	2112	17283	93083	102450	16775	927875
2	48004	242818	103414	124379	118140	52754	34231	2126322
3	1391	11779	2118	16210	21397	42701	3179	548805
4	7446	18848	6419	60994	261785	81367	100561	2503147
5	7728	13951	7546	60375	85200	24730	58757	889956
6	133560	84369	42750	167620	260158	75002	100842	4206849
7	2086	50756	7950	29523	22641	11200	2466	540274
8	21893	18402	31875	10191	11670	3137	2059	349990
9	10788	99656	15231	164890	106156	120847	25640	2007668
10	517763	26394	50459	23977	26811	14303	7563	907658
11	22491	893910	124478	78302	89627	39742	15378	1804876
12	52346	121534	694220	33487	42853	13362	6662	1183845
13	22689	74181	32266	410876	150510	66226	41365	1440092
14	24894	89400	46227	145677	821467	151255	152044	2400086
15	13294	38943	12352	66648	149602	542212	37905	1368050
16	6620	14267	5001	39229	150324	36686	491938	1084416
Sum	894911	1809347	1184418	1449661	2411424	1377974	1097365	24289909

these zones; for instance, the mean number of trips whose origin is Coyoacán (id = 2) and destination is Iztapalapa (id = 6) during a week day is 228,272 (see Table 3) and the mean number of trips whose origin is Tláhuac (id = 10) and destination is Cuauhtémoc (id = 14) is 21,881 (see Table 6) during a day on the weekend.

In Tables 3–6, the last row represents the total number of trips attracted by each zone and the last column in Tables 4 and 6 represents the total number of trips generated by each zone.

Table 5. Mean number of trips from zones 1-16 to zones 1-9 during a day of the weekend.

	1	2	3	4	5	6	7	8	9
1	331141	13688	3114	68471	5834	10607	1203	378	9658
2	11696	551750	4838	25849	19910	120115	30555	16952	52884
3	2174	4091	175522	3007	3271	7323	4482	492	65008
4	69138	24356	4142	1007667	16605	29428	825	874	11786
5	4545	20175	3729	16417	210228	128570	1118	793	8682
6	11046	124120	7325	32746	128502	1670601	4991	7800	43020
7	1311	29099	3224	1916	1200	4804	158791	3232	51829
8	252	13668	0	925	793	7594	1677	143275	2146
9	8721	52044	68219	13616	9358	40811	50897	3938	579592
10	1406	26844	757	5531	3253	100356	1831	13138	6297
11	6821	147822	2869	17403	8955	55210	32614	12158	55182
12	2622	58174	1034	4675	5281	27902	3341	25524	9751
13	10562	63766	7127	29544	39022	84688	15222	3119	85047
14	56596	66481	13774	184327	57030	180473	13321	8988	60055
15	53266	17612	17869	37333	13774	36801	4374	4100	59194
16	9800	19479	3206	72083	33526	67913	2627	3114	22571
Sum	581097	1233169	316749	1521510	556542	2573196	327869	247875	1122702

Table 6. Mean number of trips from zones 1-16 to zones 10-16 during a day of the weekend.

	10	11	12	13	14	15	16	Sum
1	1501	5786	1542	9701	63331	53846	11181	590982
2	26855	150388	60418	64149	67953	19583	19995	1243890
3	493	3714	856	8218	13863	18850	3176	314540
4	6370	16430	3851	28840	181498	37641	71621	1511072
5	3140	9292	8080	40741	56400	13737	33730	559377
6	99372	52020	30083	94294	181288	40162	68541	2595911
7	1013	37941	3235	14790	14315	4452	3892	335044
8	13264	10182	23451	2704	11629	3264	1256	236080
9	5619	54948	10135	87576	65512	60836	26414	1138236
10	345171	17204	50337	14740	21881	6563	5956	621265
11	15078	546694	77342	41336	58766	21318	10691	1110259
12	51898	76431	452836	17491	33051	8737	6028	784776
13	14753	38855	18884	232506	85671	28496	20327	777589
14	22350	60141	30310	84877	427114	92312	114306	1472455
15	7156	19593	10801	27493	94657	339935	28238	772196
16	6200	10326	5581	20049	116777	26215	344904	764371
Sum	620233	1109945	787742	789505	1493706	775947	770256	14828043

This way we can see in Table 4 that during a week day 849,911 trips are attracted to zone 10 and 540,274 trips are generated from zone 7.

To compute the new number of trips made using the subway, RTP/M1, trolleybus, light rail or suburban, we used the corresponding rates given by the government of Mexico City. To compute the number of trips made using collective/micro or buses, we used the rates of transit stations given by Google and for personal transportation we used the workplaces rate. To compute the number of trips made by bicycle, we used the rates for Ecobici and for metrobus/mexibus we used an average of both the rates of metrobus and mexibus in Mexico City. The other transport modes remain the same as in 2017.

Fig 2 shows the variations in the mobility indices from 27 February to 31 November, for each mode of transport that was modified.

The population of each borough is also considered in our mathematical model. According to [38], these populations in 2020 are given in Table 7.

Note that both, the mobility rates from Google and the government of Mexico City and the reference O-D matrix (the one from 2017) are given per day. This way, the population N_i could be considered as constant; but only per day. In this work, a phenomenon whose duration is of the order of months is studied, so that for the entire period of estimation N_i is considered as a variable.

Furthermore, although in all boroughs the largest number of trips are carried out within the same borough, the distribution of trips to/from the other boroughs does not follow the same pattern in all cases. For example, the Coyoacán borough attracts the highest number of trips from the Tlalpan borough and the least amount of trips from Cuajimalpa; meanwhile, the borough of Iztapalapa attracts the highest number of trips from the Cuauhtémoc borough and the least amount of trips from La Magdalena Contreras. This phenomenon can be explained by the prevailing economic activity in each borough and the transportation connectivity with the others.
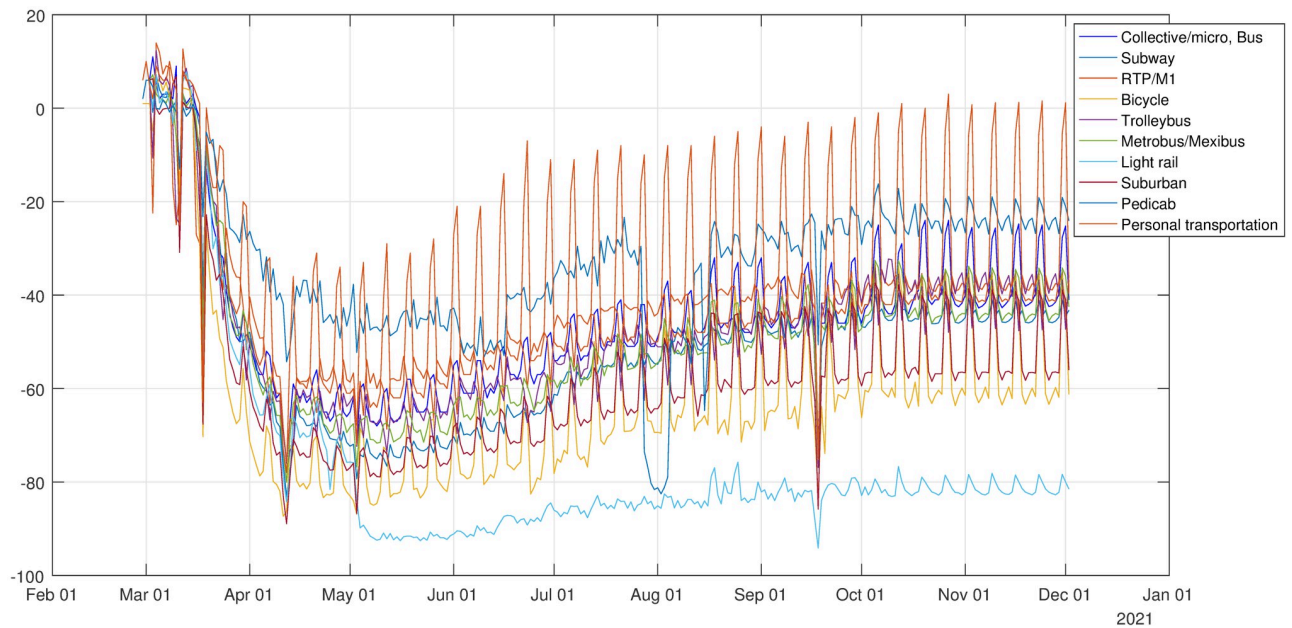
Fig 2. Variations in the mobility indices from 27 February, 2020 to 31 November, 2020.

https://doi.org/10.1371/journal.pone.0263367.g002

In order to obtain an stable algorithm for modeling correctly the migration of people, we use the Fratar method to balance all the origin-destination matrices [39].

Clusters

In this section, we describe the clustering analysis that we implemented on Mexico City based on mobility data. This analysis is presented not only to try to find some possible socioeconomic relations between some boroughs, but because there exist computational challenges

Table 7. Population in 2020 for each borough of Mexico City.

| Id	Borough	Population
1	Azcapotzalco	414,711
2	Coyoacán	628,063
3	Cuajumalpa de Morelos	199,224
4	Gustavo A. Madero	1,185,772
5	Iztacalco	384,326
6	Iztapalapa	1,815,786
7	La Magdalena Contreras	239,086
8	Milpa Alta	137,927
9	Álvaro Obreón	726,664
10	Tláhuac	305,076
11	Tlalpan	574,577
12	Xochimilco	407,885
13	Benito Juárez	385,439
14	Cuauhtémoc	531,831
15	Miguel Hidalgo	372,889
16	Venustiano Carranza	430,978

https://doi.org/10.1371/journal.pone.0263367.t007
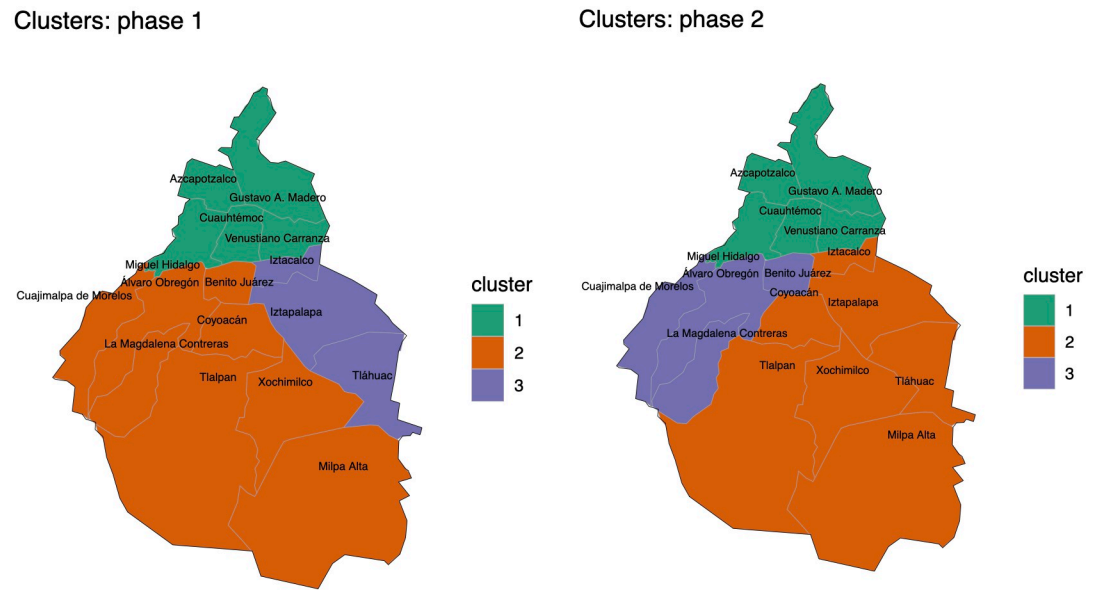
Fig 3. Borough clusters of Mexico City. (A): Borough clusters for the first period of the pandemic. (B): Borough clusters for the second period of the pandemic.

https://doi.org/10.1371/journal.pone.0263367.g003

which can be avoided creating clusters. Firstly, in order to solve model (1) for the whole Mexico City, the program uses around 50GB in RAM during the compilation process using the Stan package. Secondly, the computation time for estimating the parameters is around 3 days. We have used a computer with Ubuntu 20.04, 64 GB in RAM and 12 cores. Therefore, we propose in this section how it could be avoided both of these computational challenges. Solving model (1) for each cluster would reduce the amount of RAM memory used and the computation time. Moreover, this strategy could be implemented simultaneously using the t-walk package for example, since the t-walk package only uses one core for execution program.

During the pandemic, the Mexican government has scheduled four phases depending of level of contagion risk. These phases corresponding to the following periods: phase 1: from 27 February February to 22 March; phase 2: from 23 March to 19 April; phase 3: from 20 April to 28 June; phase 4: from 29 June to 27 October. The mobility network was analysed using the community detection module Louvain inside the *igraph* R package [40]. For more details about the *igraph* R package, see [41]. Thus, using the Louvain community detection algorithm, we are able to identify that Mexico City's network has a modular structure, with three communities, as shown in Figs 3 and 4. From Figs 3 and 4 we observe that the communities in the first and fourth phases of the pandemic are the same, and the second and third of the pandemic are the same. The community 1 of the first phase of the pandemic is composed of boroughs 1, 4, 14, 15 and 16; community 2 is composed of boroughs 2, 3, 7, 8, 9, 11, 12 and 13; and, community 3 is composed of boroughs 5, 6 and 10. The community 1 of the second phase of the pandemic is composed of boroughs 1, 4, 14, 15 and 16; community 2 is composed of boroughs 2, 5, 6, 8, 10, 11, and 12; and, community 3 is composed of boroughs 3, 7, 9 and 13.

Mathematical model

As we mention in the introduction, the transmission model incorporates information on human movement within the following Susceptible, Exposed, Infected, Recovered (SEIR)
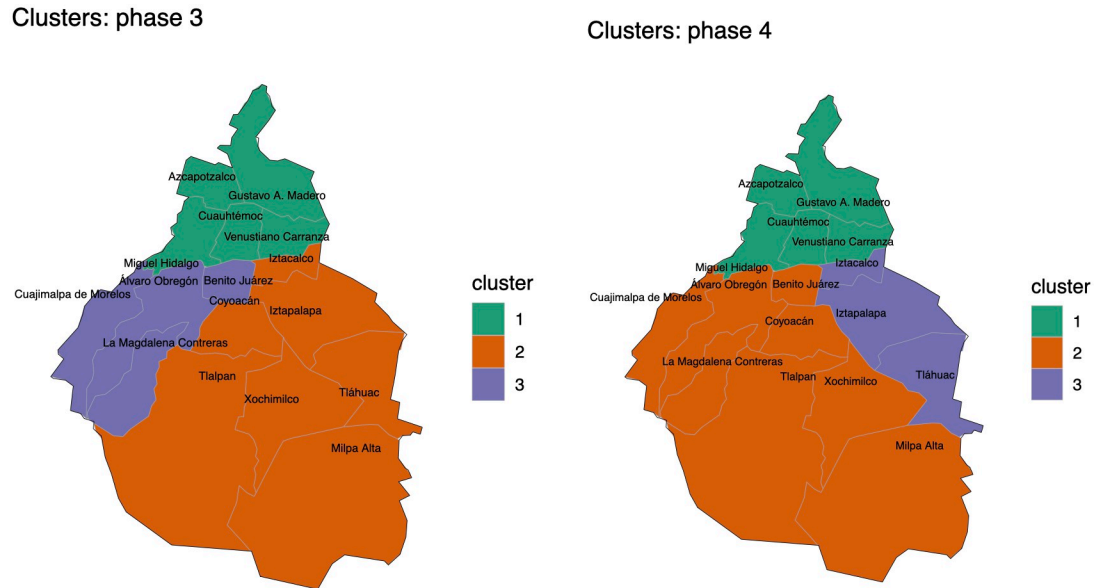
Clusters: phase 3

Clusters: phase 4

Fig 4. Borough clusters of Mexico City. (A): Borough clusters for the third period of the pandemic. (B): Borough clusters for the fourth period of the pandemic.

metapopulation structure [21]:

$$
\begin{cases}
\dfrac{dS_i}{dt} = -\dfrac{\beta_s S_i I_i}{N_i} - \dfrac{\beta_a S_i A_i}{N_i} + \theta \sum_j \dfrac{M_{ij} S_j}{N_j - I_j} - \theta \sum_j \dfrac{M_{ji} S_i}{N_i - I_i} \\[3mm]
\dfrac{dE_i}{dt} = \dfrac{\beta_s S_i I_i}{N_i} + \dfrac{\beta_a S_i A_i}{N_i} - \alpha E_i + \theta \sum_j \dfrac{M_{ij} E_j}{N_j - I_j} - \theta \sum_j \dfrac{M_{ji} E_i}{N_i - I_i} \\[3mm]
\dfrac{dA_i}{dt} = (1-\rho)\alpha E_i - \gamma A_i + \theta \sum_j \dfrac{M_{ij} A_j}{N_j - I_j} - \theta \sum_j \dfrac{M_{ji} A_i}{N_i - I_i} \\[3mm]
\dfrac{dI_i}{dt} = \rho\alpha E_i - \gamma I_i \\[3mm]
N_i = N_i^0 + \theta \sum_j M_{ij} - \theta \sum_j M_{ji}, \quad j = 1, 2, \ldots, n,
\end{cases}
\tag{1}
$$

where S_i, E_i, A_i, I_i and N_i are the susceptible, exposed, undocumented infected, documented infected and the total population in borough i at time t, respectively, and N_i^0 denotes the fixed population in borough i given by Table 7. Spatial coupling within the model is represented by the daily number of people traveling from city j to city i (M_{ij}) an a multiplicative scale factor θ, reflecting the under-reporting of human movement. It is also assumed that documented infected individuals (I_i) do not move between boroughs, although these individuals can move between boroughs during the latency period. The total population N_i in each borough is reset

Table 8. Parameter description and values proposed in [21] of the state equations given on (1).

Parameter	Description	Proposed Range	Units
β_s	Transmission rate due to documented infected individuals.	[0., 1.]	None
β_a	Transmission rate due to undocumented individuals.	[0., 1.]	None
$1/\theta$	Multiplicative factor (*reflect under-reporting of human movement*).	[0., 1.]	*Trips/person*
$1/\alpha$	Average latency period.	[2, 10]	*Days*
$1/\gamma$	Average duration of infection.	[2, 20]	*Days*
ρ	Fraction undocumented/documented people.	[0.,.4]	None

The values are given in days as time unit.

each new day as the sum of N_i^0 and the inflow term $\theta \Sigma_j M_{ij}$, minus the outflow term $\theta \Sigma_j M_{ji}$. We note that distinction between daytime and nighttime in the transmission model 1 is implemented in [35]. A complete description of the parameters involved in the model (1), the respective range of values proposed in [21] and their measurement units can be found in Table 8. We have set a minimum value for the denominator $N_j - I_j$ or $N_i - I_i$ as equal to 10^3 in order to avoid instabilities.

Parameter estimation

For parameter estimation, we use the daily reported dataset [23]. We use Bayesian inference to solve the inverse problem associated to the system of Ordinary Differential Equations (ODEs) given on (1), similarly to [33]. Some references using this method of parameter estimation can be found in [42–53].

Let us denote the vector of state variables in the zone i as $\mathbf{x} = (S_i, E_i, A_i, I_i) \in (L^2[0, T])^n$, where $n = 4$ denotes the number of state variables and the vector of parameters in the zone i as $\boldsymbol{\theta} = (\beta_s, \beta_a, \theta, \rho, \gamma, \alpha, S_{i_0}, E_{i_0}, A_{i_0}, I_{i_0}) \in \mathbb{R}^m$, where $m = 10$ denotes the dimension number of parameters to estimate. Thus, we can write the model (1) as the following Cauchy problem

$$\begin{cases} \dot{\mathbf{x}} = \varphi(\mathbf{x}, \boldsymbol{\theta}) \\ \mathbf{x}(0) = \mathbf{x}_0. \end{cases} \qquad (2)$$

Problem (2), defines a mapping $\Phi(\boldsymbol{\theta}) = \mathbf{x}$ from parameters $\boldsymbol{\theta}$ to state variables $\mathbf{x}$, where $\Phi : \mathbb{R}^m_+ \rightarrow (L^2([0, T])^n$, where $\mathbb{R}_+$ denotes the non-negative real numbers. We assume that Φ has a Fréchet derivative. Usually, not all states of the system can actually be directed measured, i.e., the data consists of measurements of some state variables at a discrete set of points $t_1, \ldots, t_k$, e.g. in epidemiology, these data consist of number of cases of confirmed infected people. This defines a linear observation mapping from state variables to data $\Psi : (L^2([0, T])^n \rightarrow \mathbb{R}^{s \times k}$, where $s \leq n$ is the number of observed variables and k is the number of sample points. Let us define $F : \mathbb{R}^m \rightarrow \mathbb{R}^{s \times k}$ as $F(\boldsymbol{\theta}) = \Psi(\Phi(\boldsymbol{\theta}))$, called the forward problem. Thus, the inverse problem is formulated as a standard optimization problem

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^m} \|F(\boldsymbol{\theta}) - \mathbf{y}_{\text{obs}}\|^2, \qquad (3)$$

such that $\mathbf{x} = \Phi(\boldsymbol{\theta})$ holds, with $\mathbf{y}_{\text{obs}}$ is the observable data which has error measurements of size $\boldsymbol{\eta}$.

Problem (3) may be solved using numerical tools to deal with a non-linear least-squares problem [54–58]. In this work, we implement Bayesian inference to solve the inverse problem

given on ([3](#)). From the Bayesian perspective, all of the state variables $\mathbf{x}$ and parameters $\boldsymbol{\theta}$ are considered as random variables and the data $\mathbf{y}_{\mathrm{obs}}$ is fixed. For the random variables $\mathbf{x}$ and $\boldsymbol{\theta}$, the joint probability distribution density of the data $\mathbf{x}$ and the parameters $\boldsymbol{\theta}$, denoted by $\pi(\boldsymbol{\theta}, \mathbf{x})$, is given by $\pi(\boldsymbol{\theta}, \mathbf{x}) = \pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$, where $\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ is the conditional probability distribution, which is also called the likelihood function, and $\pi(\mathbf{x}|\boldsymbol{\theta})$ is the prior distribution, which involves the prior information of parameters $\boldsymbol{\theta}$. Given $\mathbf{x} = \mathbf{y}_{\mathrm{obs}}$, the conditional probability distribution $\pi(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}})$, which is called the posterior distribution of $\boldsymbol{\theta}$, is given by the Bayes' theorem:

$$\pi(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}}) \propto \pi(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{4}$$

If an additive noise is assumed

$$\mathbf{y}_{\mathrm{obs}} = F(\boldsymbol{\theta}) + \boldsymbol{\eta},$$

where $\boldsymbol{\eta}$ is the noise due to discretisation, the model error and the measurement error. If the noise probability distribution $\pi_H(\boldsymbol{\eta})$ is known, $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are independent, then

$$\pi(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\theta}) = \pi_H(\mathbf{y}_{\mathrm{obs}} - F(\boldsymbol{\theta})).$$

All of the available information regarding the unknown parameter $\boldsymbol{\theta}$ is codified into a prior distribution $\pi(\boldsymbol{\theta})$, which specifies our belief in a parameter before observing the data. All of the available information regarding how we obtained the measured data is codified into the likelihood distribution $\pi(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\theta})$. This likelihood can be seen as an objective or cost function because it punishes deviations of the model from the data. To solve the associated inverse problem ([4](#)), one may use the maximum a posterior (MAP)

$$\boldsymbol{\theta}_{\mathrm{MAP}} = \max_{\boldsymbol{\theta}} \ \pi(\boldsymbol{\theta}|\mathbf{x}).$$

We used the dataset in the zone i as $\mathbf{y}_{\mathrm{obs}} = (\tilde{S}_i, \tilde{E}_i, \tilde{A}_i, \tilde{I}_i)$, which correspond to the susceptible, exposed, documented infected and undocumented infected in the zone i, respectively. A Poisson distribution with respect to the time is typically used to account for the discrete nature of these counts. However, the variance of each component of the dataset $\mathbf{y}_{\mathrm{obs}}$ is larger than its mean, which indicates that there is over-dispersion of the data. Thus, a more appropriate likelihood distribution is to use the Negative Binomial (NB) because it has an additional parameter that allows the variance to exceed the mean [50, 51, 59]. The NB is a mixture of Poisson and Gamma distributions, where the rate parameter of the Poisson distribution itself follows a Gamma distribution [59, 60]. We note that although there are different mathematical expressions for the NB depending on the author or source, they are equivalent. Because of this multiple representation of the NB in the literature, one must ensure to use the NB distribution accordingly to the source. Here, we have used the following expression for the NB distribution

$$\mathcal{NB}(y|\mu, \phi) = \frac{\Gamma(y + \phi)}{\Gamma(y)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi}\right)^y \left(\frac{\phi}{\mu + \phi}\right)^\phi, \tag{5}$$

where μ is the mean of the random variable $y \sim \mathcal{NB}(y|\mu, \phi)$ and ϕ is the over-dispersion parameter; that is,

$$\mathbb{E}[Y] \ = \mu, \qquad \mathrm{Var}(Y) \ = \mu + \frac{\mu^2}{\phi}.$$

We recall that the Poisson distribution has mean and variance equal to μ, so $\mu^2/\phi > 0$ is the additional variance of the NB with respect to the Poisson distribution. The inverse of the parameter ϕ, controls the over-dispersion. Thus, it is important to select its support adequately

for parameter estimation. In addition, there are alternative forms of the NB distribution. We have used the first option *neg_bin* of the NB distribution of Stan [61]. We acknowledge that some scientists have had success with the second alternative representation of the NB distribution [47]. We assume independent NB distributed noise $\boldsymbol{\eta}$ (i.e., all dependency in the data is codified into the contact tracing model). In other words, the positive definite noise covariance matrix $\boldsymbol{\eta}$ is assumed to be diagonal. Therefore, using the Bayes formula, the likelihood is

$$\pi(\boldsymbol{\theta}|(\tilde{I}_i) \propto \pi(\tilde{I}_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

where *i* denotes the borough index. As mentioned earlier, we approximate the likelihood probability distribution corresponding to diagnosed cases with a NB distribution

$$\tilde{I}_i^j \sim \mathcal{NB}(I_i^j(\theta), \phi_i), \tag{6}$$

where the index *j* denotes the number of days, *i* the number of the boroughs, and ϕ_i are the parameters corresponding to the over-dispersion parameter of the NB distribution (5) respect to each borough.

For independent observations, the likelihood distribution $\pi(\mathbf{y}|\boldsymbol{\theta})$ is given by the product of the individual probability densities of the observations

$$\pi(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}) = \prod_{j=1}^{n}\pi(\tilde{I}_i^j|\theta),$$

where the mean μ of the NB distribution $\mathcal{NB}(I_i(\theta), \phi_i)$, is given by the solution $I_i(t)$ of the model (1) at time $t = t_j$. For the prior distribution, we select the `LogNormal` distribution for β_s and β_a parameters, `Gamma` distributions for α and γ parameters and Uniform distributions for the other parameters to estimate: ρ, $\boldsymbol{\theta}$ and initial conditions $(S_{i_0}, E_{i_0}, A_{i_0}, I_{i_0})$. The hyperparameters and their support corresponding to all the distributions of the parameters to estimate are given on the table's range Table 8.

$$\pi(\theta) \quad = \prod_{i=1}^{n} \mathcal{LN}(a_\beta, b_\beta)\mathcal{U}(a_q, b_q)\mathcal{U}(a_\delta, b_\delta)\mathcal{U}(a_\alpha, b_\alpha)\mathcal{U}(a_\gamma, b_\gamma)$$

$$\tag{7}$$

$$\times \ \mathcal{U}(a_{s_0}, b_{s_0})\mathcal{U}(a_{E_0}, b_{E_0})\mathcal{U}(a_{I_0}, b_{I_0})\mathcal{U}(a_{Q_0}, b_{Q_0}).$$

The posterior distribution $\pi(\boldsymbol{\theta}|y_{\text{obs}})$ given by (4) does not have an analytical closed form since the likelihood function, which depends on the solution of the non-linear model given on (1), does not have an explicit solution. Then, we explore the posterior distribution using two methods: first, the Stan Statistics package [61] within its version the Automatic Differentiation Variational Inference (ADVI) method,; and second, the general purpose Markov Chain Monte Carlo Metropolis-Hasting (MCMC-MH) algorithm t- walk [62]. Both algorithms generate samples from the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}})$ that can then be used to estimate marginal posterior densities, mean, credible intervals, percentiles, variances, and others. We the reader refer to [63] for a more complex description of the MCMC-MH algorithms.

Fig 5 shows the credible intervals of parameters of model (1) within 95% Highest-Posterior Density (HPD) using the ADVI-Fullrank method of Stan package [61]. Table 9 shows the posterior mean and quantiles of all the estimated parameters of model (1) using the ADVI-Fullrank method of Stan package. Table 10 shows the posterior mean and quantiles of all over-dispersion parameters ϕ_i of the Negative Binomial distribution (6). Fig 6 shows the joint probability density distributions of the estimated parameters of model (1) within 95% (HPD). The blue lines represent the medians. Figs 7–9 show the fit of confirmed COVID-19 cases of all of
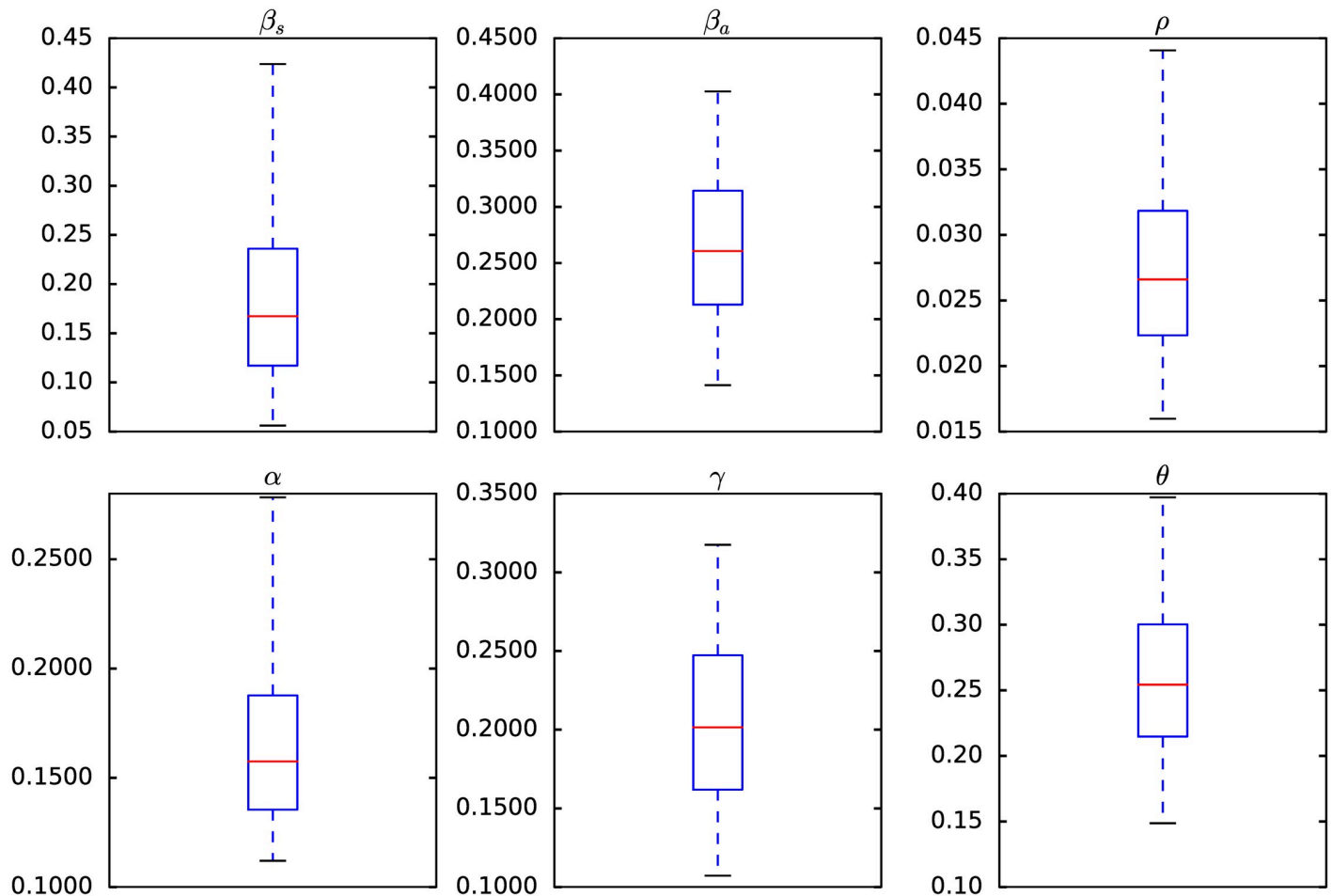
Fig 5. Credible intervals of parameters of the model (1) within 95% Highest-Posterior Density (HPD) using the Stan package [61].

the boroughs of Mexico City using the Stan [61]. Fig 10 shows Credible intervals of parameters of model (1) within 95% Highest-Posterior Density (HPD) using the t-walk Package [62]. Note that the result obtained with the t-walk package are preliminary because we only performed 60,000 iterations, with 30,000 of them as burn-in and it was obtained without balancing the Origen-Destination matrices. We performed this limited quantity of iterations because the computational time consumption is significantly large for each 1,000 of iterations. However, we will perform more iterations in the near future. Using both packages, we did a fit for the first 245 days of the pandemic in Mexico City, starting 27 February, and we have performed predictions from 245–275 days, corresponding to 28 October to 27 November and compared with the true cases in this last period 28 October to 27 November. We assumed that the mobility from 28 October to 27 November is the same as from 28 September to 27 October, i.e., we assumed the same mobility cluster for the projection period. We set up a minimum borough fraction equal to 0.6 to limit the borough to fall below their population size. Our future work will analyse the identifiability of the parameters of model (1), as suggested in [49, 64, 65]. Specifically, the ρ parameter because it is multiplied by the period of incubation of the disease, α. Thus, estimating both parameters simultaneously may lead to non-identifiability difficulty. we

Table 9. Parameter estimation of β_s, β_a, ρ, α, γ, θ and initial conditions of the model (1).

	Mean	Std	Min	2.5%	50%	97.5%	Max
β_s	0.1817	0.0688	0.0433	0.1299	0.1719	0.2215	0.4822
β_a	0.2797	0.0724	0.0903	0.2274	0.2743	0.3259	0.5665
ρ	0.0186	0.0056	0.0057	0.0146	0.0178	0.0220	0.0481
α	0.1720	0.0334	0.1103	0.1485	0.1670	0.1910	0.3067
γ	0.2210	0.0570	0.0804	0.1798	0.2161	0.2597	0.4116
θ	0.7279	0.2112	0.0409	0.6113	0.7786	0.8976	0.9963
E_{10}	215.5595	165.6988	9.4025	89.7281	172.8620	297.3270	901.8550
A_{10}	546.1551	275.5846	13.5082	310.5730	565.1880	793.2150	995.1790
I_{10}	1.9900	2.0874	0.0481	0.6862	1.3693	2.6035	22.6618
E_{20}	253.3650	193.9120	3.8161	102.6910	201.1370	366.7250	936.9310
A_{20}	295.6783	195.3005	9.7637	135.7270	260.0070	417.1160	873.4510
I_{20}	3.4961	4.5445	0.0385	0.9100	1.9425	4.0675	40.6196
E_{30}	185.7097	145.4584	1.9960	78.6950	147.7400	247.2920	835.4920
A_{30}	507.7907	263.5192	6.1474	283.0790	514.8280	736.8860	984.3130
I_{30}	4.2297	5.8422	0.0378	0.9611	2.2060	4.9589	61.0710
E_{40}	683.1201	253.4763	18.5321	517.4730	743.1590	900.7840	999.2030
A_{40}	496.4247	271.0262	6.0559	258.5220	495.7330	728.1480	987.1500
I_{40}	30.8663	29.6572	0.0269	5.2495	19.7807	52.1318	99.6383
E_{50}	552.5027	299.4180	4.4051	284.7780	577.9160	837.3670	999.1490
A_{50}	562.7039	282.2821	1.3356	315.3790	605.4310	809.2500	997.2760
I_{50}	4.9965	6.9620	0.0388	1.0097	2.4918	6.0742	61.8987
E_{60}	450.4365	279.1041	1.4010	200.7940	420.6680	688.3850	992.9970
A_{60}	547.0965	270.4777	9.5825	315.7010	568.5270	783.6100	992.2320
I_{60}	23.4015	26.0301	0.0173	3.7063	12.1013	36.0308	99.6128
E_{70}	115.0333	110.9839	2.8503	38.3297	76.6577	151.7080	795.5370
A_{70}	515.9971	278.4386	6.2676	281.7120	534.1290	755.6920	997.6560
I_{70}	3.7288	6.0063	0.0180	0.6917	1.6631	4.0766	55.2651
E_{80}	224.1849	186.9898	3.0370	75.5207	172.3130	321.0610	925.9820
A_{80}	569.0899	287.3211	1.8729	328.3040	597.3180	833.6790	999.2210
I_{80}	3.4849	4.8696	0.0357	0.7935	1.7130	4.2425	44.0485
E_{90}	188.5786	164.5613	3.3092	61.5219	134.4440	269.3350	807.6380
A_{90}	687.0370	275.9913	10.8989	500.1660	765.8410	926.9680	999.2580
I_{90}	16.5153	20.7336	0.0258	2.3711	7.4698	22.8262	96.2754
E_{100}	356.7580	263.3986	2.5233	125.7860	297.0560	550.1360	987.3250
A_{100}	746.1666	233.5140	56.2971	610.6280	829.0390	934.2400	999.0530
I_{100}	31.6034	32.8951	0.0017	3.3353	16.6582	55.7096	99.9472
E_{110}	281.1917	236.3608	2.7780	90.8408	206.4170	432.5200	970.9040
A_{110}	252.0265	216.5767	1.4617	74.8393	192.5350	375.3250	980.7710
I_{110}	4.0954	5.9613	0.0178	0.8408	1.9202	4.5660	54.8201
E_{120}	130.6564	131.6714	2.0295	36.7382	85.9181	175.2460	823.6720
A_{120}	189.1367	168.3434	3.8207	62.8391	135.1090	254.0920	847.2030
I_{120}	8.7603	12.7619	0.0338	1.1913	3.9130	10.2037	94.1881
E_{130}	135.1793	130.8547	2.4587	39.7409	89.4389	187.5800	786.0350
A_{130}	233.7060	218.5237	2.5230	63.5877	160.4860	337.5610	966.4870
I_{130}	17.2452	23.4479	0.0013	1.2447	6.5493	23.5534	99.6969
E_{140}	288.4052	241.8465	2.0949	85.1094	216.5890	441.7050	967.2260
A_{140}	132.8599	136.0876	1.5013	36.5287	86.8088	176.9500	817.9200

(Continued)

Table 9. (*Continued*)

	Mean	Std	Min	2.5%	50%	97.5%	Max
I_{140}	1.1226	1.1440	0.0290	0.4021	0.7689	1.4526	13.1217
E_{150}	194.1867	170.3316	2.6126	63.5594	138.1430	273.0570	818.0280
A_{150}	629.8427	301.4854	5.1016	382.9950	700.1620	909.0470	999.4200
I_{150}	31.9147	32.6907	0.0114	3.2822	18.5268	56.7282	99.9608
E_{160}	615.1693	319.6856	0.5475	323.0720	703.2440	916.7530	999.8430
A_{160}	590.1725	297.5308	7.9860	345.1590	634.7790	868.4580	999.4370
I_{160}	34.3333	34.6819	0.0100	3.8419	18.8559	64.8812	99.9809

have uploaded all the codes and source data used in this paper to the following Github link for a detailed review.

The basic reproduction number estimation

The basic reproduction number, which is commonly denoted by $\mathcal{R}_0$, is the average number of secondary infections generated by a single infective during the curse of the infection in a whole susceptible population. We calculate the reproductive number $\mathcal{R}_0$ in Mexico City using the inferred parameters. Define $\mathbf{X} = (E, A, I)$ and using the next generation operator method [66] on the system (1), the Jacobian matrices $\mathcal{F}$ and $\mathcal{V}$ of system (1) are given by

$$
\mathcal{F} = \begin{pmatrix} \dfrac{\beta_s SA}{N} + \dfrac{\beta_a SI}{N} \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathcal{V} = \begin{pmatrix} \alpha E \\ \gamma A - \rho \alpha E \\ \gamma I - (1-\rho)\alpha E \end{pmatrix}.
$$

Table 10. Over-dispersion parameters estimation of the Negative Binomial distribution (6).

|	Mean	Std	Min	2.5%	50%	97.5%	Max
ϕ_1	2.1991	0.3290	1.2822	1.9601	2.1692	2.4130	3.5188
ϕ_2	1.8943	0.3639	1.0883	1.6340	1.8487	2.1194	3.9229
ϕ_3	2.0723	0.3572	1.2465	1.8068	2.0520	2.2942	3.2018
ϕ_4	1.4408	0.2664	0.8347	1.2476	1.4203	1.6004	3.0505
ϕ_5	1.8577	0.2999	1.1289	1.6330	1.8409	2.0425	2.9154
ϕ_6	1.2810	0.2120	0.6839	1.1406	1.2595	1.4009	2.1667
ϕ_7	2.0110	0.3294	1.0595	1.7742	1.9821	2.2169	3.3500
ϕ_8	1.5453	0.3538	0.7403	1.2915	1.5145	1.7583	3.3937
ϕ_9	2.1633	0.4467	1.1252	1.8410	2.1092	2.4173	4.3067
ϕ_{10}	1.3869	0.2924	0.6578	1.1816	1.3610	1.5593	2.4075
ϕ_{11}	2.4184	0.4398	1.1707	2.1076	2.3709	2.6832	4.0595
ϕ_{12}	1.8298	0.4035	0.8828	1.5461	1.7847	2.0668	4.1388
ϕ_{13}	2.6204	0.6058	1.2449	2.2169	2.5623	2.9725	6.2347
ϕ_{14}	2.1933	0.3804	1.3056	1.9111	2.1554	2.4154	3.8624
ϕ_{15}	2.1977	0.4987	1.0955	1.8362	2.1534	2.5010	3.9972
ϕ_{16}	1.6167	0.3232	0.8505	1.3926	1.5958	1.8085	3.0021
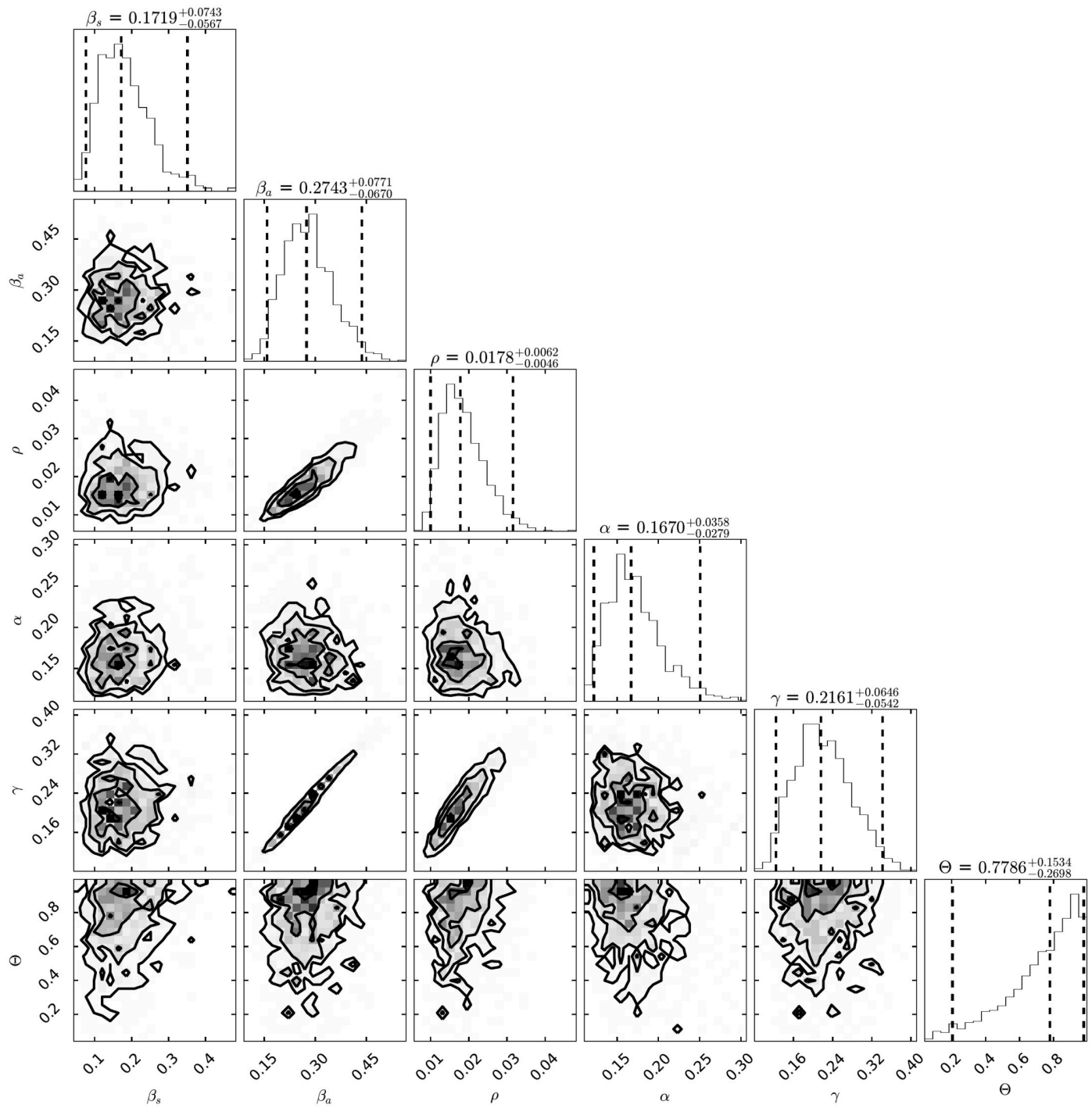
Fig 6. Joint probability density distributions of the estimated parameters of model (1) within 95% (HPD). The blue lines represent the medians.
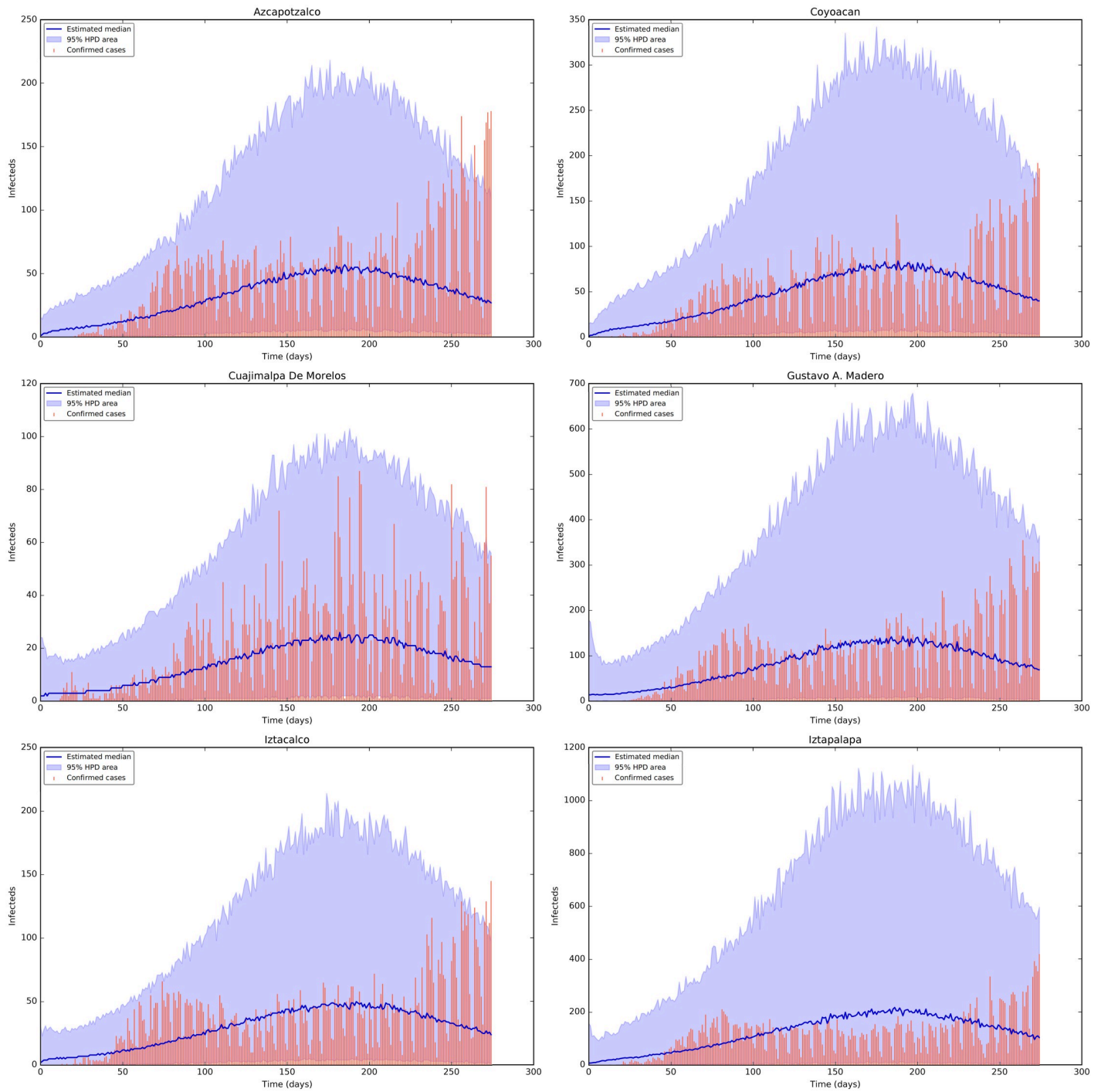
Fig 7. Fit of confirmed COVID-19 cases of the boroughs 1 to 6 using the Stan package [61]. Top row from left-hand to right-hand: the fit for the confirmed cases of the Districts Azcapotzalco and Coyoacan. The tomato colour bars represent the confirmed cases, the blue and purple solid lines represent the median and the mode, respectively, and the shaded area represent the %95 probability bands for the expected value for the state variable of Documented Infecteds. Middle row from left-hand to right-hand: the fit for the diagnosed cases of the Districts Cuajimalpa de Morelos and Gustavo A. Madero. Bottom row from left-hand to right-hand: the fit for the diagnosed cases of the Districts Iztacalco and Iztapalapa.
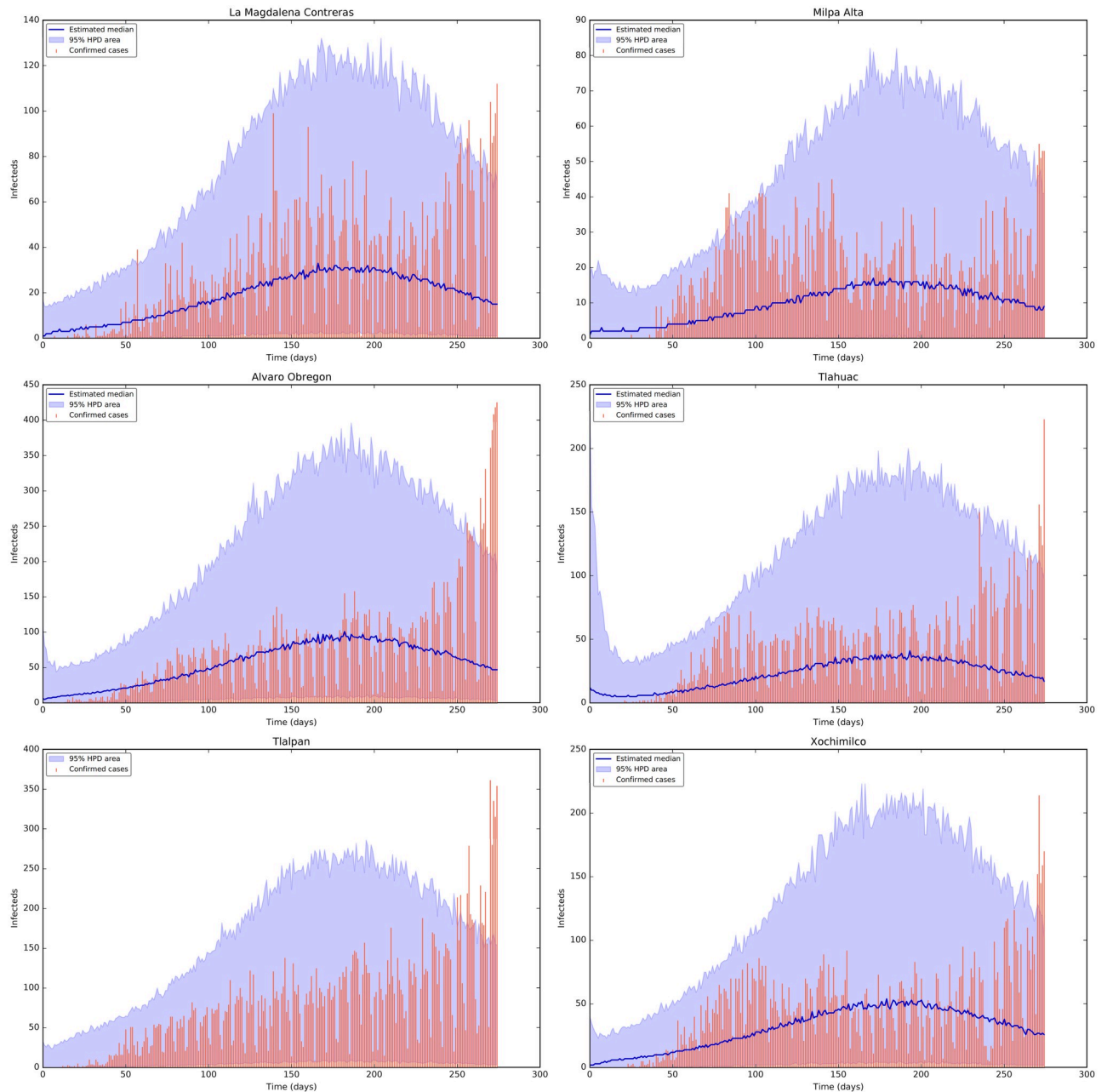
https://doi.org/10.1371/journal.pone.0263367.g007

Fig 8. Fit of confirmed COVID-19 cases of the boroughs 7 to 12 using the Stan package [61]. Top row from left-hand to right-hand: the fit for the confirmed COVID-19 cases of the Districts La Magdalena Contreras and Milpa Alta. The tomato colour bars represent the confirmed COVID-19 cases, the blue solid line represent the median and the shaded area represent the %95 probability bands for the expected value for the state variable of Documented Infecteds. Middle row from left-hand to right-hand: the fit for the diagnosed cases of the Districts Alvaro Obregon and Tlahuac. Bottom row from left-hand to right-hand: the fit for the diagnosed cases of the Districts Tlalpan and Xochimilco.
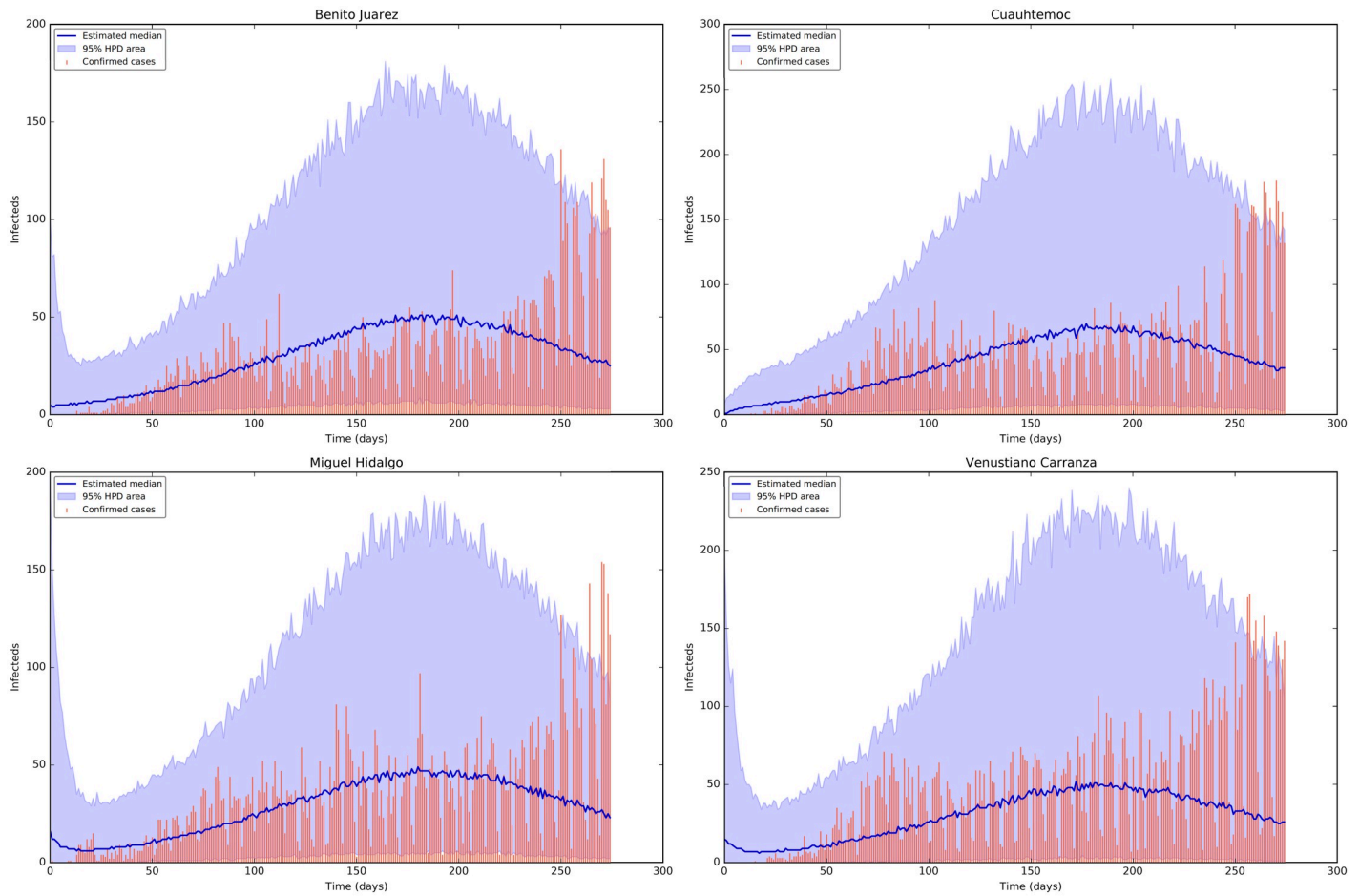
https://doi.org/10.1371/journal.pone.0263367.g008

Fig 9. Fit of confirmed COVID-19 cases of the boroughs 13 to 16 using the Stan package [61]. Top row from left-hand to right-hand: the fit for the confirmed COVID-19 cases of the Districts Benito Juarez and Cuauhtemoc. The tomato colour bars reprent the confirmed COVID-19 cases, the blue solid line reprent the median and the shaded area represent the %95 probability bands for the expected value for the state variable of Documented Infecteds. Bottom row from left-hand to right-hand: the fit for the diagnosed cases of the Districts Miguel Hidalgo and Venustiano Carranza.

The disease free equilibrium (DFE) of system (1) is $\mathbf{X}_0 = (0, 0, 0, N, 0)^T$, we then have

$$\mathbf{F} = \frac{\partial \mathcal{F}}{d\mathbf{X}}(\mathbf{X}_0) = \begin{pmatrix} 0 & \beta_s & \beta_a \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$\mathbf{V} = \frac{\partial \mathcal{V}}{d\mathbf{X}}(\mathbf{X}_0) \begin{pmatrix} \alpha & 0 & 0 \\ -\rho\alpha & \gamma & 0 \\ -(1-\rho)\alpha & 0 & \gamma \end{pmatrix}.$$

Therefore, the next-generation matrix is $\mathbf{K} = \mathbf{FV}^{-1}$, from where R_e is computed as the leading
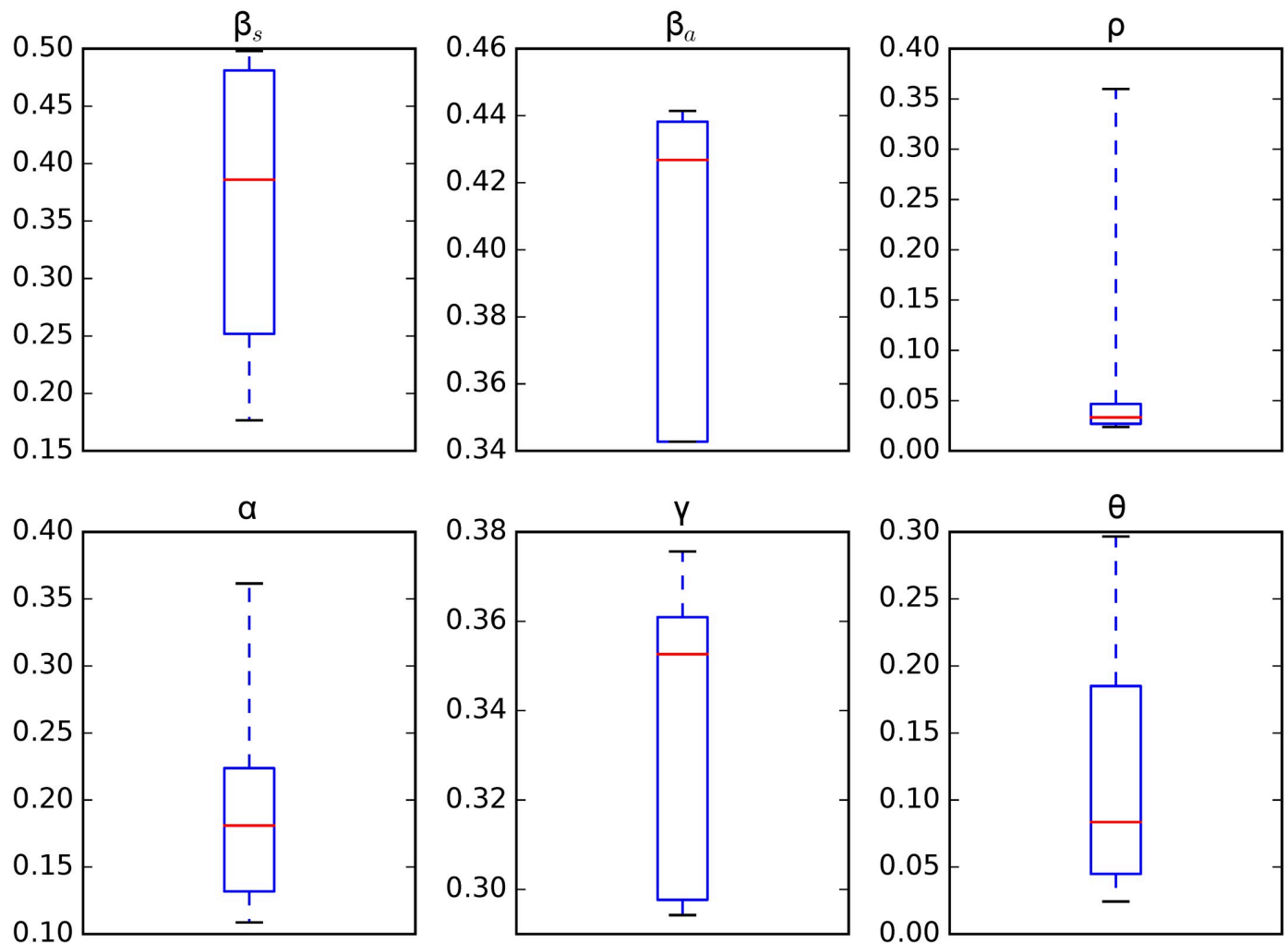
Fig 10. Credible intervals of parameters of model (1) within 95% Highest-Posterior Density (HPD) using the t-walk package [62].

eigenvalue of matrix **K**; that is,

$$\mathcal{R}_0 = \frac{\rho \beta_s}{\gamma} + \frac{(1 - \rho)\beta_a}{\gamma}.$$ (8)

Table 8 shows the range of values for the parameters involved in the expression (8) obtained using the Stan package [61]. With those values, we obtain a %95 credible interval for $\mathcal{R}_0 \in (1.0224, 1.2376)$.

Discussion

In this work, we analyse a networked dynamic metapopulation model of the coronavirus dissemination in Mexico City using ODEs and Bayesian statistics. We present an explanation of how to estimate the mobility per day between the boroughs that compound the Mexico City, both on a weekday and on weekends; combining available information from the origin-destination survey carried out in 2017 with the current mobility indices that Google and the

government of Mexico City report, depending on the mode of transport used to make each trip (e.g., bus, subway, car, etc.) and then we use the Fratar method to balance the daily origin-destination matrices. We also present a clustering analysis of the boroughs which compound Mexico City based on mobility data from Google and the Transportation Mexican Survey. From Figs 3 and 4, we can identify three different clusters during the each phase of the pandemic. We point out that the same cluster analysis done for Mexico City, could be implemented for a broader area, the metropolitan area named Valle de Mexico, which is rather important for the whole country. We consider that this clustering analysis which is based on individual movement may be crucial to efficiently model a human pandemic on the same scale as presented here, or at a smaller scale.

From Fig 5, the transmission rate of symptomatic was 0.19 within 95% Credible Interval (CI) [0.06, 0.42], and the transmission rate of asymptomatic was 0.27 within 95% CI [0.14, 0.40], which is in concordance with the value estimated of 0.25 in [67], in that study the transmission rate was not separated in symptomatics and asyntomatics. The fraction of undocumented infections, ρ, was 0.027 within 95% CI [0.02, 0.04]. The estimated latency period, $1/\alpha$, is 5.96 days within 95% CI [3.60, 8.93] days, which is in concordance with the value used of 5.99, 6.0, 5.1 and 5.0 in [50, 52, 68, 69], and the estimated recovery period, $1/\gamma$, is 4.86 days within 95% CI [3.15, 9.33] days, which is lower in comparison with the ones used of 5.97 and 10.81 (asymptomatic and symptomatic class, respectively) in [69], 10.0 in [52], 7.0 in [50] and 5.0, 10.8 and 14.0 (reported infectious, symptomatic and asymptomatic class, respectively) in [68]. The inter-borough scale factor θ was 0.77 within 95% CI [0.61, 0.89], this value indicates that the mean number of trips made by a person is between three and four in one day, which makes sense with complete trips to get out of home, do some activities (e.g., work, shopping, or services), and return home. The results of the inferred parameters of model (1) and the population size of the boroughs (e.g., Iztapalapa and Gustavo A. Madero) help to explain the fast dispersion of COVID-19 and indicate the challenge of finding strategies to contain it. We have compared the parameter values inferred with respect to those used for Mexico City. As mentioned in Section, we will analyse the identifiability of the parameters of model (1) (i.e., the ρ parameter) because this parameter is multiplied by the period of incubation of the disease, α. Thus, estimating both parameters simultaneously may lead to a non-identifiability difficulty. We may observe this non-identifiability in Figs 5 and 10; that is, different combinations of the model parameters lead to the same "energy" value of the system 1. In particular, we can observe that a different combination of the estimated parameters values obtained with the methods ADVI-Fullrank and ADVI-Meanfield give very similar fitted curves for diagnosed cases. We also observe that the recovery period time is more in accordance with the values used in [50, 52, 68, 69] but the latency period is lower than the ones used there. We note that the parameters, β_s, β_a of the model (1) are considered as global; that is, they are assumed the same for all the boroughs of Mexico City and all the transportation modes. In the near future, we will explore a more robust model that will consist of local parameters of transmission β_s, β_a, instead of globally (i.e., a pair of transmission rates β_s, β_a for each borough). Furthermore, we will consider those transmission rates, β_s, β_a, to be dependent on time as in [52, 70]. In addition, we will consider the interstate and international mobility from/to Mexico City. We will take into account imported cases from the other 31 states of Mexico. We will also consider the cases imported from overseas by airplane passengers, and will do a global and local sensitivity analysis of model (1). Finally, we will investigate a spatio-temporal model based on a diffusion partial differential equation model combined with individual movement trends.

Author Contributions

Conceptualization: Kernel Prieto, M. Victoria Chávez–Hernández.

Formal analysis: Kernel Prieto, M. Victoria Chávez–Hernández.

Investigation: Kernel Prieto, Jhoana P. Romero–Leiton.

Methodology: Kernel Prieto, M. Victoria Chávez–Hernández, Jhoana P. Romero–Leiton.

Resources: Kernel Prieto.

Software: Kernel Prieto, M. Victoria Chávez–Hernández.

Validation: Kernel Prieto.

Writing – original draft: Kernel Prieto.

Writing – review & editing: Kernel Prieto, Jhoana P. Romero–Leiton.

References

1. World Health Organization. Novel Coronavirus (2019-nCoV) Situation Reports; 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation.

2. Paules CI, Marston HD, Fauci AS. Coronavirus infections more than just the common cold. Jama. 2020; 323(8):707–708. https://doi.org/10.1001/jama.2020.0757 PMID: 31971553

3. Izquierdo LD, et al. Informe técnico nuevo coronavirus 2019-nCoV. Instituto de Salud Carlos III; 2020. Available from: https://www.mscbs.gob.es/va/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/20200210_ITCoronavirus.pdf.

4. Chen TM, Rui J, Wang QP, Zhao ZY, Cui JA, Yin L. A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. Infectious Diseases of Poverty. 2020; 9(1):1–8. https://doi.org/10.1186/s40249-020-00640-3 PMID: 32111262

5. Choi S, Ki M. Estimating the reproductive number and the outbreak size of novel coronavirus disease (COVID-19) using mathematical model in Republic of Korea. Epidemiology and Health. 2020; p. 123–145.

6. Yang C, Wang J. A mathematical model for the novel coronavirus epidemic in Wuhan, China. Mathematical Biosciences and Engineering. 2020; 17(3):2708–2724. https://doi.org/10.3934/mbe.2020148 PMID: 32233562

7. Savi PV, Savi MA, Borges B. A Mathematical Description of the Dynamics of Coronavirus Disease (COVID-10): A Case Study of Brazil. arXiv preprint arXiv:200403495. 2020.

8. Shaikh AS, Shaikh IN, Nisar KS. A Mathematical Model of COVID-19 Using Fractional Derivative: Outbreak in India with Dynamics of Transmission and Control. Preprints. 2020.

9. Yang W, Zhang D, Peng L, Zhuge C, Hong L. Rational evaluation of various epidemic models based on the COVID-19 data of China. arXiv preprint arXiv:200305666. 2020.

10. Nesteruk I. Statistics-based predictions of coronavirus epidemic spreading in mainland China. medRxiv. 2020.

11. Zhong L, Mu L, Li J, Wang J, Yin Z, Liu D. Early prediction of the 2019 novel coronavirus outbreak in the mainland china based on simple mathematical model. Ieee Access. 2020; 8:51761–51769. https://doi.org/10.1109/ACCESS.2020.2979599 PMID: 32391240

12. Zhao Z, Zhu YZ, Xu JW, Hu QQ, Lei Z, Rui J, et al. A mathematical model for estimating the age-specific transmissibility of a novel coronavirus. medRxiv. 2020.

13. Miller A, Foti N, Lewnard J, Jewell N, Guestrin C, Fox E. Mobility trends provide a leading indicator of changes in SARS-CoV-2 transmission. medRxiv. 2020.

14. Pei S, Shaman J. Initial simulation of SARS-CoV2 spread and intervention effects in the continental US. medRxiv. 2020.

15. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2011.

16. Keeling M, Rohani P. Estimating spatial coupling in epidemiological systems: a mechanistic approach. Ecology Letters. 2002; 5:20–29. https://doi.org/10.1046/j.1461-0248.2002.00268.x

17. Ma Z. Spatiotemporal fluctuation scaling law and metapopulation modeling of the novel coronavirus (COVID-19) and SARS outbreaks. arXiv preprint arXiv:200303714. 2020.

18. Calvetti D, Hoover A, Rose J, Somersalo E. Metapopulation network models for understanding, predicting and managing the coronavirus disease COVID-19. arXiv preprint arXiv:200506137. 2020.

19. Wells K, Lurgi M. COVID-19 containment policies through time may cost more lives at metapopulation level. medRxiv. 2020.

20. Coletti P, Libin P, Petrof O, Willem L, Steven A, Herzog SA, et al. A data-driven metapopulation model for the Belgian COVID-19 epidemic: assessing the impact of lockdown and exit strategies. medRxiv. 2020.

21. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science. 2020; 368(6490):489–493. https://doi.org/10.1126/science.abb3221 PMID: 32179701

22. Zhou C. Evaluating new evidence in the early dynamics of the novel coronavirus COVID-19 outbreak in Wuhan, China with real time domestic traffic and potential asymptomatic transmissions. medRxiv. 2020.

23. de Mexico G. Covid-19 Mexico; 2020. https://coronavirus.gob.mx/datos/.

24. INEGI. Encuesta Origen Destino en Hogares de la Zona Metropolitana del Valle de México (EOD) 2017; 2017. https://www.inegi.org.mx/programas/eod/2017/.

25. Google. Informes de movilidad local sobre el COVID-19; 2020. https://www.google.com/covid19/mobility/.

26. de la Ciudad de México G. Afluencia preliminar en transporte público; 2020. https://datos.cdmx.gob.mx/dataset/afluencia-preliminar-en-transporte-publico.

27. McBryde E, Gibson G, Pettitt A, Zhang Y, Zhao B, McElwain D. Bayesian modelling of an epidemic of severe acute respiratory syndrome. Bulletin of mathematical biology. 2006; 68(4):889–917. https://doi.org/10.1007/s11538-005-9005-4 PMID: 16802088

28. Nour M, Cömert Z, Polat K. A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization. Applied Soft Computing. 2020; p. 106580. https://doi.org/10.1016/j.asoc.2020.106580 PMID: 32837453

29. Şenel K, Özdinç M, Öztürkcan DS, Akgül A. Instantaneous R for COVID-19 in Turkey: estimation by Bayesian statistical inference [Türkiye'de COVID-19 için anlık R hesaplaması: Bayesyen istatistiksel çıkarım ile tahmin]. Turkiye Klinikleri Journal of Medical Sciences. 2020; 40(2):127–131.

30. Bagal DK, Rath A, Barua A, Patnaik D. Estimating the parameters of susceptible-infected-recovered model of COVID-19 cases in India during lockdown periods. Chaos, Solitons & Fractals. 2020; 140:110154. https://doi.org/10.1016/j.chaos.2020.110154 PMID: 32834642

31. Liu F, Li X, Zhu G. Using the contact network model and Metropolis-Hastings sampling to reconstruct the COVID-19 spread on the "Diamond Princess". Science Bulletin. 2020;. https://doi.org/10.1016/j.scib.2020.04.043 PMID: 32373394

32. Batistela CM, Correa DP, Buenoc ÁM, Piqueira JR. Compartmental model with loss of immunity: analysis and parameters estimation for Covid-19. arXiv preprint arXiv:200701295. 2020.

33. Prieto K. Current forecast of COVID-19: Bayesian and Machine Learning approaches. medRxiv. 2020.

34. Worldometer. Worldometer Coronavirus Updates; 2020. https://www.worldometers.info/.

35. Pei S, Kandula S, Yang W, Shaman J. Forecasting the spatial transmission of influenza in the United States. Proceedings of the National Academy of Sciences. 2018; 115(11):2752–2757. https://doi.org/10.1073/pnas.1708856115 PMID: 29483256

36. Chávez Hernández MV, Juárez Valencia LH, Ríos Solís YA. Penalization and augmented Lagrangian for O-D Demand Matrix Estimation from Transit Segment Counts. Transportmetrica A: Transport Science. 2019; 15(2):915–943.

37. Spiess H. A maximum likelihood model for estimating origin-destination matrices. Transportation Research Part B: Methodological. 1987; 21(5):395–412. https://doi.org/10.1016/0191-2615(87)90037-3

38. INEGI. Instituto Nacional de Estadística y Geografía; 2020. https://www.inegi.org.mx/temas/estructura/default.html#Publicaciones/.

39. Chávez Hernández M. V. (2014). Modelos matemáticos para análisis de demanda en transporte. *Mathematics Department*, *UAM-Iztapalapa*.

40. de Anda-Jáuregui G. COVID-19 in Mexico: A network of epidemics. arXiv e-prints. 2020.

41. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal. 2006; Complex Systems:1695.

42. Stojanović O, Leugering J, Pipa G, Ghozzi S, Ullrich A. A Bayesian Monte Monte Carlo approach for predicting the spread of infectious diseases. PLoS ONE. 2019; 14(12). https://doi.org/10.1371/journal.pone.0225838 PMID: 31851680

43. Luzyanina T, Bocharov G. Markov chain Monte Carlo parameter estimation of the ODE compartmental cell growth model. Mathematical Biology and Bioinformatics. 2018; 13:376–391. https://doi.org/10. 17537/2018.13.376

44. Brown G, Porter A, Oleson J, Hinman J. Approximate Bayesian computation for spatial SEIR(S) epidemic models. Spatial and Spatio'temporal Epidemiology. 2018; 24(10):2685–2697. https://doi.org/10. 1016/j.sste.2017.11.001 PMID: 29413712

45. Bettencourt L, Ribeiro R. Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases. PlosOne. 2008; 3(5):e2185. https://doi.org/10.1371/journal.pone.0002185 PMID: 18478118

46. Boersch-Supan P, Ryan S, Johnson L. deBInfer:Bayesian inference for dynamical models of biological systems in R. Methods in Ecology and Evolution. 2017; 8:511–518. https://doi.org/10.1111/2041-210X. 12679

47. Grinsztajn L, Semenova E, Margossian C, Riou J. Bayesian workflow for disease transmission modeling in Stan. arXiv e-prints. 2020.

48. Bliznashki S. A Bayesian Logistic Growth Model for the Spread of COVID-19 in New York. medRxiv. 2020; 14(12).

49. Chowell G. Fitting dynamic models to epidemic outbreak with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. Infectious Disease Modelling. 2017; 2:379–398. https://doi.org/10.1016/j.idm.2017.08.001 PMID: 29250607

50. Capistrán M, Capella A, Christen A. Forecasting hospital demand during COVID-19 pandemic outbreaks. arXiv e-prints. 2020.

51. Argüedas Y, Santana-Cibrian M, Velasco-Hernández J. Transmission dynamics of acute respiratory diseases in a population structured by age. Mathematical Biosciences and Engineering. 2019; 16 (6):7477–7493. https://doi.org/10.3934/mbe.2019375 PMID: 31698624

52. Acuña Zegarra M, Comas-García A, Hernández-Vargas E, Santana-Cibrian M, Velasco-Hernández J. The SARS-CoV-2 epidemic outbreak: a review of plausible scenarios of containment and mitigation for Mexico. medRxiv. 2020.

53. Chatzilena A, Leeuwen E, Ratmann O, Baguelin M, Demiris N. Contemporary statistical inference for infectious disease models using Stan. Epidemics. 2019; 29. https://doi.org/10.1016/j.epidem.2019. 100367 PMID: 31591003

54. Prieto K and Dorn O. Sparsity and level set regularization for diffuse optical tomography using a transport model in 2D. Inverse Problems. 2016; 33(1):014001. https://doi.org/10.1088/0266-5611/33/1/ 014001

55. Prieto K and Ibarguen-Mondragon E. Parameter estimation, sensitivity and control strategies analysis in the spread of influenza in Mexico. Journal of Physics: Conference Series. 2019; 1408(1):012020.

56. Smirnova A, DeCamp L, Liu H. In: Inverse problems and ebola virus disease using an age of infection model. Springer, Cham; 2016. p. 103–121.

57. Alavez-Ramirez J. Estimacion de parámetros en ecuaciones diferenciales ordinarias: identificabilidad y aplicaciones a medicina. Revista electrónica de contenido matemático. 2007; 21.

58. Capistrán M, Moreles M, Lara B. Parameter estimation of some epidemic models: The case of recurrent epidemics caused by respiratory syncytial virus. Bulletin of Mathematical Biology. 2009; 71(8):1890–1901. https://doi.org/10.1007/s11538-009-9429-3 PMID: 19568727

59. Nayens T and Faes C and Molenberghs G. A generalized Poisson-gamma model for spatially oversdispersed data. Spatial and Spatio-temporal Epidemiology. 2012; 3:185–194. https://doi.org/10.1016/j. sste.2011.10.004

60. Coly S, Yao A, Abrial D, Garrido M. Disributions to model overdispersed count data. Journal de la Societe Francaise de Statistique. 2016; 157(2):39–63.

61. Carpenter B, Gelman A, Hoffman D, Goodrich B, Betancourt M, Brubaker M, et al. Stan: A probabilistic programming language. Journal of Statistical Software. 2017; 76(1):1–32. https://doi.org/10.18637/jss. v076.i01

62. Christen J, Fox C. A general purpose sampling algorithm for continuous distributions (the t-walk). Bayesian Anal. 2010; 5:263–282. https://doi.org/10.1214/10-BA603

63. House T, Ford A, Lan S, Bilson S, Buckingham-Jeffery E, Girolami M. Bayesian uncertainty quantification for transmissibility of influenza, norovirus and Ebola using information geometry. Journal of the Royal Society Interface. 2016; 13. https://doi.org/10.1098/rsif.2016.0279 PMID: 27558850

64. Roosa K, Chowell G. Assesing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. Theoretical Biology and Medical Modelling. 2019; 16(1). https://doi.org/10.1186/s12976-018-0097-6 PMID: 30642334

65. Magal P, Webb G. The parameter identification problem for SIR epidemic models: Identifying unreported cases. Journal of Mathematical Biology. 2018; 77:1629–1648. https://doi.org/10.1007/s00285-017-1203-9 PMID: 29330615

66. Diekmann O, Heesterbeek JAP, Metz JA. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. Journal of mathematical biology. 1990; 28(4):365–382. https://doi.org/10.1007/BF00178324 PMID: 2117040

67. Dbouk T, Drikakis D. Fluid dynamics and epidemiology: Seasonality and transmission dynamics. Physics of Fluids. 2021; 33(2):021901. https://doi.org/10.1063/5.0037640 PMID: 33746486

68. Saldaña F, Velasco-Hernández J. The trade-off between mobility and vaccination for COVID-19 control: a metapopulation modeling approach. medRxiv. 2020.

69. Acuña Zegarra M, Santana-Cibrian M, Velasco-Hernández J. Modeling behavioral change and COVID-19 containment in Mexico: A trade-off between lockdown and compliance. Mathematical Biosciences. 2020; 325. https://doi.org/10.1016/j.mbs.2020.108370 PMID: 32387384

70. Piccolomini E, Zama F. Monitoring Italian COVID-19 spread by a forced SEIRD model. PLoS ONE. 2020; 15(8):e0237417. https://doi.org/10.1371/journal.pone.0237417