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

Kernel Prieto¹
M. Victoria Chávez-Hernández²
Jhoana P. RomeroLeiton³

¹Instituto de Matemáticas, Universidad Autónoma de México, México

²Facultad de Ingeniería Mecánica y Eléctrica, Universidad Autónoma de Nuevo León, México

³Facultad de Ingeniería, Universidad Cesmag, Colombia

¹Autor Corresponsal email: kernel@ciencias.unam.mx

²Email id: vicky.fis@gmail.com

³Email id: jpatirom3@gmail.com

January 24, 2021

Abstract

This work presents a forecast of the spread of the new coronavirus in Mexico City based on a mathematical model with metapopulation structure by using Bayesian Statistics inspired in a data-driven approach. The mobility of humans on a daily basis in Mexico City is mathematically represented by a origin-destination matrix using the open mobility data from Google and a Transportation Mexican Survey. This matrix, is incorporated in a compartmental model. We calibrate the model against borough-level incidence data collected between February 27, 2020 and October 27, 2020 using Bayesian inference to estimate critical epidemiological characteristics associated with the coronavirus spread. Since working with metapopulation models lead to rather high computational time consume, we do a clustering analysis based on mobility trends in order to work on these clusters of borough separately instead of taken all the boroughs together at once. This clustering analysis could be implemented in smaller or larger scale in different part of the world. In addition, this clustering analysis is divided in the phases that the government of Mexico City has set up to restrict the individuals movement in the city. Also, we calculate the reproductive number in Mexico City using the next generation operator method and the inferred model parameters. The analysis of mobility trends can be helpful in public health decisions.

Keywords: Bayesian inference, COVID-19, metapopulation model, mobility trends.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

1 Introduction

Coronavirus disease 2019 (COVID-19) is an illness caused by a novel coronavirus. Coronaviruses are a family of viruses that cause infection in humans and animals. Diseases by coronavirus are zoonotic [58]. Coronaviruses that affect humans (HCoV) can produce clinical symptoms including Severe Acute Respiratory Syndrome (SARS) viruses and Middle East Respiratory Syndrome (MERS-CoV) [45]. This disease 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 December 31, 2019. On March 11, 2020, the WHO declared COVID-19 a global pandemic [32]. Since the beginning of the epidemic, to January 21, 2021 have been reported more than 97,890,676 cases and more than 2,094,459 deaths in the world.

The first case of COVID-19 in South America was registered in Brazil on February 26, 2020. The first death from this infection in this region was announced in Argentina on March 7, 2020. Then, the virus arrived to Mexico, where it is now becoming a large problem with almost 1,688,944 confirmed cases and 144,371 deaths to January 21, 2021. To date, many researchers around the world have focused their interests on understanding the transmission dynamics of COVID-19 disease using mathematical and statistical models and methods, see e.g., [16, 17, 43, 51, 53, 60, 61, 62, 63], but we will focus on those that incorporate information on human movement. A relation between human mobility and transmission of coronavirus disease in USA has been studied in [41, 46]. Metapopulation models are one of the simplest spatial models, but they are also one of the most applicable to modelling many human diseases [34]. 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 [33] and recently to model the spread of COVID-19 see e.g., [10, 20, 35, 38, 57].

In this work, we calibrate the metapopulation model proposed by Li et. al. in [35], similar to [64], using incidence data reported in [25]. For this end, we first describe the mathematical model used, then we computed the number of trips that are produced and attracted in some borough of Mexico City using data about these trips in 2017 [30] and combining them with the rates of reduction or increase in mobility during the pandemic reported by Google [27] and the government of Mexico City [24]. Then, by using Bayesian inference we solve the associated inverse problem to predict the dynamics of the spread of cases similarly as in the references [4, 5, 36, 40, 44, 47, 52]. Conclusions are presented in the last section.

2 Computation of the mobility matrices

In order to incorporate the mobility in the transmission model, the produced and attracted trips in the boroughs of Mexico City are considered (see Figure 1 and Table 1). Mexico City is the capital of Mexico, with around 9 million inhabitants and floating population of over 22 million composed of daily commuters and international visitors. Mexico city belongs to the top ten most crowded cities in the world [59]. It has a large number of headquarters and a large transport network, composed of 20 different modes of transport.

Figure 1: Municipalities 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
4	Gustavo A. Madero	8	Milpa Alta	12	Xochimilco		
				13	Benito Juárez		

Table 1: Municipalities of Mexico City.

The mobility between the zones in Table 1 is represented in a two-dimensional arrangement known as the origin-destination matrix (O-D matrix) $\mathbf{M} = \{M_{ij}\}$, $i, j = 1, \dots, 9$, 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 one was carried out in 2017 [30]. 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. In this paper, we consider all the trips and are identified by the mode of transport that was used to carry out in the area of interest (see Table 2).

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

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

To update the O-D matrix, there are several methodologies 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 routes [15]. There are also approaches that project the trips to/from each zone based on the projected economic growth in those areas [55]. Nevertheless, given the pandemic situation 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 update it to a scenario in 2020 using the daily mobility reports provided by Google [27] and the government of Mexico City [24]. According to [30], tables 3 to 6 represent the number of trips between these zones; for instance, the mean number of trips which 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 which origin is Tláhuac (id = 10) and destination is Cuauhtémoc (id = 14) is 21,881 (see Table 6) during a day of the weekend.

Table 3: Mean number of trips from zones 1-16 to zones 1-8 during a week day.

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

Table 4: Mean number of trips from zones 1-16 to zones 9-16 during a week day.

	9	10	11	12	13	14	15	16
1	24149	1918	10139	2112	17283	93083	102450	16775
2	108741	48004	242818	103414	124379	118140	52754	34231
3	100180	1391	11779	2118	16210	21397	42701	3179
4	27846	7446	18848	6419	60994	261785	81367	100561
5	19190	7728	13951	7546	60375	85200	24730	58757
6	64521	133560	84369	42750	167620	260158	75002	100842
7	83675	2086	50756	7950	29523	22641	11200	2466
8	3901	21893	18402	31875	10191	11670	3137	2059
9	1034673	10788	99656	15231	164890	106156	120847	25640
10	11070	517763	26394	50459	23977	26811	14303	7563
11	95685	22491	893910	124478	78302	89627	39742	15378
12	14993	52346	121534	694220	33487	42853	13362	6662
13	161543	22689	74181	32266	410876	150510	66226	41365
14	107543	24894	89400	46227	145677	821467	151255	152044
15	115902	13294	38943	12352	66648	149602	542212	37905
16	22583	6620	14267	5001	39229	150324	36686	491938

Table 5: Mean number of trips from zones 1-16 to zones 1-8 during a day of the weekend.

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

Table 6: Mean number of trips from zones 1-16 to zones 9-16 during a day of the weekend.

	9	10	11	12	13	14	15	16
1	9658	1501	5786	1542	9701	63331	53846	11181
2	52884	26855	150388	60418	64149	67953	19583	19995
3	65008	493	3714	856	8218	13863	18850	3176
4	11786	6370	16430	3851	28840	181498	37641	71621
5	8682	3140	9292	8080	40741	56400	13737	33730
6	43020	99372	52020	30083	94294	181288	40162	68541
7	51829	1013	37941	3235	14790	14315	4452	3892
8	2146	13264	10182	23451	2704	11629	3264	1256
9	579592	5619	54948	10135	87576	65512	60836	26414
10	6297	345171	17204	50337	14740	21881	6563	5956
11	55182	15078	546694	77342	41336	58766	21318	10691
12	9751	51898	76431	452836	17491	33051	8737	6028
13	85047	14753	38855	18884	232506	85671	28496	20327
14	60055	22350	60141	30310	84877	427114	92312	114306
15	59194	7156	19593	10801	27493	94657	339935	28238
16	22571	6200	10326	5581	20049	116777	26215	344904

To compute the new number of trips done using 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 done using collective/micro or buses, we used the rates of transit stations given by Google and for personal transportation we used the workplaces rate. In the case of the trips done 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 rest of the transport modes remain the same as in 2017.

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

Figure 2: Boroughs of Mexico City.

In our mathematical model, the population of each borough is also considered. According to [31], those populations in 2020 are given in Table 7.

Table 7: Population in 2020 for each municipality of Mexico City.

Id	Municipality	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 Obregón	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

2.1 Clusters

In this section, we describe the clustering analysis implemented on Mexico City based on mobility data. 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 February 27th February to 22nd March; phase 2: from March 23rd to April 19th; phase 3: from April 20th to June 28th; phase 4: from June 29th to October 27th. The mobility network was analysed using the community detection module Louvain inside the *igraph* R package [23]. For more details about the *igraph* R package, see [22]. Thus, using the Louvain community detection algorithm, we identify that the Mexico City network has a modular structure, with 3 communities as shown in 3. These communities are the following: left top row shows the first phase of the pandemic. Community 1 is composed by boroughs 2, 3, 7, 8, 9, 11, 12 and 13, community 2 is composed by boroughs 1, 4, 15 and 16 and community 3 is composed by boroughs 5, 6 and 10. Right top row shows the second phase of the pandemic. Community 1 is composed by boroughs 3, 7, 9 and 13, community 2 is composed by boroughs 1, 4, 14, 15 and 16 and community 3 is composed by boroughs 2, 5, 6, 8, 10, 11 and 12. Left bottom row shows the third phase of the pandemic. Community 1 is composed by boroughs 3, 7, 9 and 13, community 2 is composed by boroughs 1, 4, 15 and 16 and community 3 is composed by boroughs 2, 5, 6, 8, 10, 11 and 12. Right bottom row shows the fourth phase of the pandemic. Community 1 is composed by boroughs 2, 3, 7, 8, 9, 11, 12 and 13, community 2 is composed by boroughs 1, 4, 15 and 16 and community 3 is composed by boroughs 5, 6 and 10. We also have done an clustering analysis for a bigger region, the Metropolitan Area of Valley of Mexico (MAVM), the most affected region of diagnosed cases of coronavirus in Mexico country in order to explore which boroughs of Mexico City share the same community with municipalities of Mexico State. The MAVM consists of all the 16 boroughs of Mexico City, 59 municipalities of the State of Mexico and 1 municipality of the state of Hidalgo. Here, we have taken into account the eighteen most representative municipalities of Mexico State. Thus, we considered a sub-area of MAVM consisting of a

total of 34 zones. The municipalities considered of Mexico State have the following identification Id: Coacalco de Berriozabal, Cuautitlan, Chalco, Chicoloapan, Chimalhuacan, Ecatepec de Morelos, Huixquilucan, Ixtapaluca, Naucalpan de Juarez, Nezahualcoyotl, Nicolas Romero, La Paz, Tecamac, Texcoco, Tlalnepantla de Baz, Tultitlan, Cuautitlan Izcalli, Balle de Chalco Solidaridad, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, and 34, respectively. Figure 4 shows that the sub-area of MAVM is divided in four communities based on mobility data. These communities remain the same during the four phases. The communities are the following: community 1 is composed by boroughs 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13 and 28, community 2 is composed by boroughs 1, 15, 17, 18, 23, 25, 27, 31, 32 and 33, community 3 is composed by boroughs 4, 14, 16, 20, 21, 22, 26, 29 and 30 and finally community 4 is composed by boroughs 19, 29 and 34. Figure 4 shows that the boroughs 1, 4, 14, 15, and 16 of Mexico City belong to the communities 2 and 3 which are not part of the main community of the Mexico City, i.e., boroughs 1 and 15 belong to the community 2 and boroughs 4, 14 and 16 belongs to the community 3. Therefore, based on the origin-destination matrices, we can see that the boroughs 1, 4, 14, 15 and 16 of Mexico City are more communicated with the Mexico State area. This could be explained in part geographically since the boroughs 1, 4, 14, 15 and 16 are in the north of Mexico City and particularly the boroughs 1, 4, 15 and 16 are on the boundary between Mexico City and Mexico State, secondly, the explanation has to do that people from Mexico State travel on daily basis to these boroughs 4, 14, 15 and 16 to work there since these boroughs concentrate business centres and corporation headquarters. From figure 4, we also can observe that the borough 28 (La Paz) from Mexico State is more connected with the Mexico City area, this may be due to the fact that it borders to the west and south-west with one of the most populated municipalities in Mexico City, Iztapalapa, and that it connects directly with three subway lines.

Figure 3: District clusters of Mexico City. Top row from left to right: the District clusters for the first period of the pandemic, and the Districts clusters of the second period of the pandemic. Bottom row from left to right: the District clusters of the third and the fourth periods of the pandemic

Figure 4: District clusters of Valle de Mexico. Top row from left to right: the District clusters of the first period of the pandemic, and the Districts clusters of the second period of the pandemic. Bottom row from left to right: the District clusters of the third and the fourth periods of the pandemic

3 Mathematical model

As we mention in the Introduction, the transmission model incorporates information on human movement within the following Susceptible, Exposed, Infected, Recovered (SEIR)

metapopulation structure [35]:

$$\left\{ \begin{array}{l} \dot{S}_i = -\frac{\beta_s S_i I_i}{N_i} - \frac{\beta_a S_i A_i}{N_i} + \theta \sum_j \frac{M_{ij} S_j}{N_j - I_j} - \theta \sum_j \frac{M_{ji} S_i}{N_i - I_i} \\ \dot{E}_i = \frac{\beta_s S_i I_i}{N_i} + \frac{\beta_a S_i A_i}{N_i} - \alpha E_i + \theta \sum_j \frac{M_{ij} E_j}{N_j - I_j} - \theta \sum_j \frac{M_{ji} E_i}{N_i - I_i} \\ \dot{A}_i = (1 - \rho) \alpha E_i - \gamma A_i + \theta \sum_j \frac{M_{ij} A_j}{N_j - I_j} - \theta \sum_j \frac{M_{ji} A_i}{N_i - I_i} \\ \dot{I}_i = \rho \alpha E_i - \gamma I_i \\ (S_i(0), E_i(0), A_i(0), I_i(0)) = (S_{i0}, E_{i0}, A_{i0}, I_{i0}) \\ N_i = N_i + \theta \sum_j M_{ij} - \theta \sum_j M_{ji}, \quad j = 1, 2, \dots, n. \end{array} \right. \quad (1)$$

In above model the symbol $\dot{\cdot}$ means the derivative in time. Additionally, S_i , E_i , I_i^r and I_i^u are the susceptible, exposed, documented infected, undocumented infected being N_i the total population in city i at time t . Also, it is assumed that individuals in the I_i^r group do not move between cities, though these individuals can move between cities during the latency period.

A complete description of the parameters involved in the model (1), the respective range of values proposed in [35] and their measurement units can be found on Table 8.

β_s	Transmission rate due to documented infected individuals.	[0., 1.]	None
β_a	Transmission rate due to undocumented individuals.	[0., 1.]	None
$\frac{1}{\theta}$	Multiplicative factor (<i>reflect under-reporting of human movement</i>).	[0., 1.]	<i>trips/person</i>
$\frac{1}{\alpha}$	Average latency period.	[2, 10]	<i>days</i>
$\frac{1}{\gamma}$	Average duration of infection.	[2, 20]	<i>days</i>
ρ	Fraction undocumented/documented people.	[0., .4]	None

Table 8: ModelParameters.

4 The basic reproduction number estimation

The basic reproduction number, commonly denoted by $\mathcal{R}_0$ is the average number of secondary infections generated by a single infective during the course of the infection in a whole susceptible population. We calculate the reproductive number R_e in Mexico City using the inferred parameters. Define $\mathbf{X} = (E, A, I)$ and using the next generation operator method [26] on the system (1), the Jacobian matrices $\mathcal{F}$ and $\mathcal{V}$ of system (1) are given by

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

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

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

and

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

Therefore, the next-generation matrix is $\mathbf{K} = \mathbf{FV}^{-1}$, from where R_e is computed as the leading eigenvalue of matrix $\mathbf{K}$. i.e,

$$\mathcal{R}_e = \frac{\rho\beta_s}{\gamma} + \frac{(1-\rho)\beta_a}{\gamma}. \quad (2)$$

Note that at the beginning of the epidemic (e.g before February 27, 2020), $\mathcal{R}_e$ is equivalent to $\mathcal{R}_0$. As the epidemic unfolds, a declination of susceptible population reduces the effective reproductive number. Here, we used the same formula above for periods after February 27, 2020.

Table 8 shows the range of values for the parameters involved in the expression (2) obtained using the Stan package [13]. With those values we obtain a %95 credible interval for $R_e \in (1.2713, 1.3054)$.

5 Parameter estimation

For parameter estimation we use the daily reported data-set [25]. For this end, we use Bayesian inference to solve the inverse problem associated to the system of Ordinary Differential Equations (ODEs) given on (1) similarly to [47]. Some references using this method of parameter estimation can be found in [1, 3, 6, 7, 8, 9, 11, 14, 18, 28, 37, 56].

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

$$\begin{cases} \dot{\mathbf{x}} = \varphi(\mathbf{x}, \boldsymbol{\theta}) \\ \mathbf{x}(0) = \mathbf{x}_0. \end{cases} \quad (3)$$

Problem (3), 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. Therefore, the inverse problem is formulated as a standard optimization problem

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^m} \|F(\boldsymbol{\theta}) - \mathbf{y}_{\text{obs}}\|^2, \quad (4)$$

such that $x = \Phi(\boldsymbol{\theta})$ holds, with $\mathbf{y}_{\text{obs}}$ is the data which has error measurements of size η . Problem (3) may be solved using numerical tools to deal with a non-linear least-squares problem [2, 12, 48, 49, 54]. In this work, we implement Bayesian inference to solve the inverse problem given on (4). From the Bayesian perspective, all state variables $\mathbf{x}$ and

parameters θ are considered as random variables and the data $\mathbf{y}_{\text{obs}}$ is fixed. For the random variables $\mathbf{x}$ and θ , the joint probability distribution density of the data $\mathbf{x}$ and the parameters θ , denoted by $\pi(\theta, \mathbf{x})$, is given by $\pi(\theta, \mathbf{x}) = \pi(\mathbf{x}|\theta)\pi(\theta)$, where $\pi(\mathbf{x}|\theta)\pi(\theta)$ is the conditional probability distribution, also called the likelihood function, and $\pi(\mathbf{x}|\theta)$ is the prior distribution which involves the prior information of parameters θ . Given $\mathbf{x} = \mathbf{y}_{\text{obs}}$, the conditional probability distribution $\pi(\theta|\mathbf{y}_{\text{obs}})$, called the posterior distribution of θ is given by the Bayes' theorem:

$$\pi(\theta|\mathbf{y}_{\text{obs}}) \propto \pi(\mathbf{y}_{\text{obs}}|\theta)\pi(\theta), \quad (5)$$

If an additive noise is assumed

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

where η is the noise due to discretization, the model error and the measurement error. If the noise probability distribution $\pi_H(\eta)$ is known, θ and η are independent, then

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

All the available information regarding the unknown parameter θ is codified into the a prior distribution $\pi(\theta)$, it specifies our belief in a parameter before observing the data. All the available information regarding the way of how was obtained the measured data is codified into the likelihood distribution $\pi(\mathbf{y}_{\text{obs}}|\theta)$. This likelihood can be seen as an objective or cost function, as it punishes deviations of the model from the data. To solve the associated inverse problem (5), one may use the maximum a posterior (MAP)

$$\theta_{\text{MAP}} = \max_{\theta} \pi(\theta|\mathbf{x}), \quad \theta_{\text{CM}} = \mathbb{E}[\pi(\theta|\mathbf{x})].$$

We used the data set in the zone i as $\mathbf{y}_{\text{obs}} = (\tilde{S}_i, \tilde{E}_i, \tilde{I}_i^r, \tilde{I}_i^u)$, 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 data set $\mathbf{y}_{\text{obs}}$ 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), since it has an additional parameter that allows the variance to exceed the mean [3, 11, 42]. In fact the NB is a mixture of Poisson and Gamma distributions, where the rate parameter of the Poisson distribution itself follows a Gamma distribution [21, 42]. We mention there exists 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 assure to use the NB distribution accordingly to the source. Here, we have used the following expression for the NB distribution

$$\mathcal{NB}(y|\mu, \phi) = \frac{\Gamma(y + \phi)}{\Gamma(y)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^y \left(\frac{\phi}{\mu + \phi} \right)^{\phi}, \quad (6)$$

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

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

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 selecting its support adequately for parameter estimation. Also, there exists alternative forms of the NB distribution. In fact, we have used the first option *neg_bin* of the NB distribution of Stan [13]. We acknowledge that some scientists have had success with the second alternative representation of the NB distribution [28]. We assume independent NB distributed noise η , i.e., all dependency in the data is codified into the contact tracing model. In other words, the positive definite noise covariance matrix η is assumed to be diagonal. Therefore, using the Bayes formula, the likelihood is

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

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

$$\tilde{I}_i^j \sim \mathcal{NB}(I_i^j(\theta), \phi_i), \quad (7)$$

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 (6) respect to each borough. For independent observations, the likelihood distribution $\pi(\mathbf{y}|\theta)$, is given by the product of the individual probability densities of the observations

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

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 rest of parameters to estimate: ρ, θ 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 table's range 8.

$$\begin{aligned} \pi(\theta) = & \prod_{i=1}^n \mathcal{LN}(a_\beta, b_\beta) \mathcal{U}(a_q, b_q) \mathcal{U}(a_\delta, b_\delta) \mathcal{U}(a_\alpha, b_\alpha) \mathcal{U}(a_\gamma, b_\gamma) \\ & \times \mathcal{U}(a_{s_0}, b_{s_0}) \mathcal{U}(a_{E_0}, b_{E_0}) \mathcal{U}(a_{I_0}, b_{I_0}) \mathcal{U}(a_{Q_0}, b_{Q_0}). \end{aligned} \quad (8)$$

The posterior distribution $\pi(\theta|y_{\text{obs}})$ given by (5) 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: the Stan Statistics package [13] within its version the Automatic Differentiation Variational Inference (ADVI) method, and secondly, the general purpose Markov Chain Monte Carlo Metropolis-Hasting (MCMC-MH) algorithm **t-walk** [19]. Both algorithms generate samples from the posterior distribution $\pi(\theta|\mathbf{y}_{\text{obs}})$ that can be used to estimate marginal posterior densities, mean, credible intervals, percentiles, variances, and others. We refer to [29] for a more complex MCMC-MH algorithms.

Figure 5 shows the credible intervals of parameters of model (1) within 95% Highest-Posterior Density (HPD) using the Stan Package [13]. Figures 6-8 show the fit of confirmed COVID-19 cases of all the boroughs of Mexico City using the Stan [13]. Figure 10 shows Credible intervals of parameters of model (1) within 95% Highest-Posterior Density

(HPD) using the ADVI-Meanfield method of Stan package [13]. Figure 11 shows Credible intervals of parameters of model (1) within 95% Highest-Posterior Density (HPD) using the t-walk Package [19]. We point out that results obtained with the t-walk package are preliminary since we performed only 60,000 iterations with 30,000 of them as burn-in. We performed this limited quantity of iterations since the computational time consumption is significantly large for each 1,000 of iterations. 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 the 27th of February, and we have performed predictions from 245 – 275 day, corresponding to October 28-th to November 28-th. We set up an minimum borough fraction equal to 0.6 to limit the borough to fall below their population size. Some future work will correspond to analyse the identifiability of the parameters of model (1), as suggested in [18, 39, 50], specifically the ρ parameter since this parameter is multiplied by the period of incubation of the disease, α , thus, estimating both parameters simultaneously may lead to non-identifiability difficulty.

Figure 5: Credible intervals of parameters of model (1) within 95% Highest-Posterior Density (HPD) using the Stan Package [13]

Figure 6: Fit of confirmed COVID-19 cases using the Stan package [13]. Top row from left to right: the fit for the confirmed cases of the Districts Azcapotzalco and Coyoacan. The tomatoe colour bars repret the confirmed cases, the blue and purple solid lines repret 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 to right: the fit for the diagnosed cases of the Districts Cuajimalpa de Morelos and Gustavo A. Madero. Bottom row from left to right: the fit for the diagnosed cases of the Districts Iztacalco and Iztapalapa.

Figure 7: Fit of confirmed COVID-19 cases using the Stan package [13]. Top row from left to right: the fit for the confirmed COVID-19 cases of the Districts La Magdalena Contreras and Milpa Alta. The tomatoe colour bars repret the confirmed COVID-19 cases, the blue solid line repret 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 to right: the fit for the diagnosed cases of the Districts Alvaro Obregon and Tlahuac. Bottom row from left to right: the fit for the diagnosed cases of the Districts Tlalpan and Xochimilco.

Figure 8: Fit of confirmed COVID-19 cases using the Stan package [13]. Top row from left to right: the fit for the confirmed COVID-19 cases of the Districts Benito Juárez and Cuauhtémoc. The tomatoe 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. Bottom row from left to right: the fit for the diagnosed cases of the Districts Miguel Hidalgo and Venustiano Carranza.

Figure 9: Joint probability density distributions of the estimated parameters within 95% (HPD). The blue lines represent the medians.

	mean	std	min	2.5%	50%	97.5%	max
β_s	0.1874	0.0970	0.0252	0.0561	0.1674	0.4237	0.5859
β_a	0.2653	0.0714	0.0837	0.1413	0.2607	0.4027	0.5472
ρ	0.0275	0.0073	0.0108	0.0160	0.0266	0.0441	0.0563
α	0.1675	0.0436	0.1023	0.1120	0.1575	0.2783	0.3721
γ	0.2057	0.0568	0.0761	0.1072	0.2015	0.3175	0.3923
θ	0.2607	0.0644	0.0995	0.1487	0.2543	0.3972	0.4845
E_{10}	184.3217	146.0019	3.3820	20.2310	143.0350	555.9260	882.5690
A_{10}	535.6228	262.4267	18.3910	58.9131	544.1240	945.0320	996.0400
I_{10}	3.3211	4.1348	0.0710	0.2166	2.0098	14.4367	54.8012
E_{20}	205.7978	169.3601	5.8993	19.8599	160.9300	643.1670	889.7030
A_{20}	209.4122	170.9916	2.4569	16.9377	159.3900	654.5110	900.7740
I_{20}	2.9190	3.1590	0.0865	0.3008	1.9768	11.1929	31.2103
E_{30}	145.8133	135.1257	2.7717	12.4263	98.9493	520.7570	777.5560
A_{30}	402.3670	245.8086	6.1237	40.3436	375.2480	894.0940	990.3010
I_{30}	4.3542	5.5809	0.0540	0.2546	2.5353	19.0474	55.5592
E_{40}	702.9435	273.8195	5.7849	86.0434	794.9580	991.9220	999.3340
A_{40}	432.1857	273.8040	10.5007	42.6838	403.6190	913.8580	990.9040
I_{40}	35.8966	30.9604	0.0317	0.5399	27.3064	95.5256	99.6371
E_{50}	537.4395	276.9246	7.5696	55.8129	551.4940	961.3700	999.5670
A_{50}	464.0044	282.1865	0.7563	34.5782	432.4700	956.6170	990.0100
I_{50}	5.4617	6.9190	0.0645	0.2225	2.8948	25.1745	51.6870
E_{60}	449.9065	295.9580	0.3066	23.2522	415.6940	956.0590	996.7500
A_{60}	432.2875	289.1433	2.7484	24.1510	395.6120	962.4810	995.5330
I_{60}	26.2795	27.2295	0.0364	0.3178	14.9401	90.2088	98.8098
E_{70}	83.9107	79.6407	2.3274	8.9936	60.3275	296.4430	643.2260
A_{70}	476.9134	267.1927	1.8749	57.2647	468.6710	941.6410	989.7840
I_{70}	2.8198	3.5169	0.0306	0.1553	1.6862	13.0140	47.4353
E_{80}	190.3153	165.8991	2.2265	16.1677	138.7190	632.7900	907.4100
A_{80}	513.3394	300.4417	1.3576	27.0607	523.6620	970.2510	997.5210
I_{80}	3.0580	5.3752	0.0239	0.1848	1.4966	13.3490	87.2188
E_{90}	180.8683	163.3549	2.7683	11.5561	130.9020	613.8010	903.1110
A_{90}	689.9932	264.9430	0.9175	112.3040	773.1550	989.1490	998.4900
I_{90}	16.5314	19.5683	0.0135	0.2542	8.4786	70.1495	98.5526
E_{100}	298.3463	237.1295	2.6486	17.2592	232.3220	837.6510	955.7100
A_{100}	623.9403	319.4492	1.7870	24.2471	713.8970	994.4600	999.7380
I_{100}	36.2087	34.8954	0.0002	0.1429	21.8228	98.7000	99.9149
E_{110}	243.8098	214.7284	2.0004	11.5777	175.9570	784.3620	968.7030
A_{110}	198.8057	180.8828	2.1306	12.4892	137.6340	679.0910	910.9480
I_{110}	5.1452	7.4224	0.0287	0.1621	2.4138	27.1594	61.6416
E_{120}	97.3070	100.6553	1.5153	7.4905	63.6840	383.0570	734.7550
A_{120}	156.2853	174.5743	0.3561	4.7232	85.3196	642.8160	972.1780
I_{120}	6.7454	9.5071	0.0206	0.1810	3.2779	35.0406	79.3457
E_{130}	103.1659	114.1921	2.0040	5.4339	60.5820	436.4240	803.5820
A_{130}	149.8156	146.1935	1.6622	8.8127	97.8790	582.2150	775.5250
I_{130}	17.5729	23.3707	0.0052	0.1023	6.2487	84.1743	99.4204
E_{140}	218.2782	210.3152	1.4959	7.4696	141.5680	773.9480	955.9370
A_{140}	99.6862	121.2945	1.2440	4.4774	51.7817	478.1060	705.4500
I_{140}	1.7250	2.3882	0.0168	0.1084	0.9497	8.2495	30.9248
E_{150}	153.4127	149.2736	1.4227	9.8697	101.7880	544.8440	826.6820
A_{150}	557.2493	333.0496	0.5477	14.0250	599.8180	993.3000	999.0760
I_{150}	36.5668	34.1058	0.0039	0.1421	25.0996	99.2558	99.9876
E_{160}	559.8946	344.4055	0.4518	8.6621	617.1160	997.2990	999.8600
A_{160}	457.4253	338.2471	0.1368	5.4778	428.0480	985.6480	999.3850
I_{160}	36.5718	32.8368	0.0210	0.3451	26.4229	97.0763	99.6811

Table 9: Parameter estimation of $\beta_s, \beta_a, \rho, \alpha, \gamma, \theta$ and initial conditions of the model (1).

	mean	std	min	2.5%	50%	97.5%	max
ϕ_1	1.8436	0.3002	1.0792	1.3056	1.8291	2.5001	3.1128
ϕ_2	1.6759	0.2968	0.8999	1.1622	1.6467	2.2845	2.9861
ϕ_3	1.7297	0.3035	0.9696	1.2452	1.6951	2.4110	3.2856
ϕ_4	1.3431	0.2549	0.6379	0.9309	1.3037	1.8718	2.5684
ϕ_5	1.6994	0.2757	0.9961	1.1949	1.6978	2.2357	2.7680
ϕ_6	1.1833	0.2087	0.6738	0.8241	1.1683	1.6235	2.0631
ϕ_7	1.6808	0.3299	0.9035	1.1439	1.6388	2.4835	3.4941
ϕ_8	1.3720	0.3556	0.5456	0.8031	1.3296	2.2064	3.3737
ϕ_9	1.8103	0.4010	0.8205	1.1819	1.7761	2.6899	3.5184
ϕ_{10}	1.2417	0.2775	0.6432	0.7851	1.2081	1.8786	2.3848
ϕ_{11}	1.9574	0.3915	0.8413	1.2730	1.9163	2.7961	3.7190
ϕ_{12}	1.6432	0.3214	0.7780	1.1183	1.6148	2.3350	2.9464
ϕ_{13}	2.3849	0.6096	1.1438	1.4321	2.3020	3.8546	5.9349
ϕ_{14}	2.0084	0.3884	1.0244	1.3742	1.9686	2.8769	4.2106
ϕ_{15}	1.9921	0.4525	0.8746	1.2461	1.9412	2.9943	3.9610
ϕ_{16}	1.4280	0.2972	0.6955	0.9395	1.4029	2.0896	2.8242

Table 10: Over-dispersion parameters' estimation of the Negative Binomial Distribution (7).

Figure 10: Credible intervals of parameters of model (1) within 95% Highest-Posterior Density (HPD) ADVI-Meanfield method of Stan package [13]

Figure 11: Credible intervals of parameters of model (1) within 95% Highest-Posterior Density (HPD) using the t-walk package [19]

6 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 between the boroughs that compound the Mexico City area both on a weekday and on weekends; we combine information available 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 (bus, subway, car, etc.). Also, we present a clustering analysis of the boroughs which compound Mexico City based on mobility data from Google and a Transportation Mexican Survey. From Figure 3, we may identify three different clusters during the each phase of the pandemic. Also, from Figure 4, we may identify four clusters during the four phases of the pandemic. We consider this clustering analysis based on individuals movement may be crucial to model efficiently a human pandemic.

From Figure 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], the fraction of undocumented infections, ρ , was 0.027 within 95% CI [0.02, 0.04]. The latency and recovery periods, $1/\alpha$ and $1/\gamma$, were 5.96 days within 95% CI [3.60, 8.93] days, and 4.86 days within 95% CI [3.15, 9.33] days respectively. The inter-borough scale factor θ was 0.26 within 95% CI [0.15, 0.40], this value indicate that the mean number of trips done by a person is between 3 and 4 in one day, which makes sense with complete trips to get out of home, do some activities (work, shopping, services), and return home. These results of the inferred parameters of model (1) and the population size of boroughs, such as, Iztapalapa and Gustavo A. Madero, explain the fast dispersion

of COVID-19 and indicate the challenge of finding strategies to contain it. As mentioned in Section 5, we will analyse the identifiability of the parameters of model (1), that is, the ρ parameter since this parameter is multiplied by the period of incubation of the disease, α , thus, estimating both parameters simultaneously may lead to non-identifiability difficulty. We may observe this non-identifiability in figures 5, 10 and 11, i.e., different combinations of the model parameters lead to the same “energy” value of the system 1. We point out that the parameters, β_s, β_a of the model (1) are considered as global, i.e., they are assumed the same for all the boroughs of Mexico City and all the transportation modes. We will explore in a near future a more robust model which will consider not only 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, but also the interstate and international mobility from/to Mexico City. We will take into account imported cases from the rest of the 30 states of the Mexico country. Also, consider the imported cases from overseas coming by airplane passengers. In addition, we will do a global and local sensitivity analysis of model (1). Also, investigate a spatio-temporal model based on a diffusion partial differential equation model combined with individual movement trends.

Acknowledgments

K. Prieto wishes to acknowledge the financial support from CONACYT through its program “Cátedras CONACYT”. M.V. Chávez-Hernández wishes to acknowledge the grant FC-2016/1948 from CONACYT, 283 Mexico. Jhoana Romero thanks the support of Fundación Ceiba, Colombia.

References

- [1] Acuña Zegarra, M., Comas-García, A., Hernández-Vargas, E., Santana-Cibrian, M., and Velasco-Hernández, J. (2020). The SARS-CoV-2 epidemic outbreak: a review of plausible scenarios of containment and mitigation for Mexico. *medRxiv*. 5
- [2] Alavez-Ramirez, J. (2007). Estimacion de parámetros en ecuaciones diferenciales ordinarias: identificabilidad y aplicaciones a medicina. *Revista electrónica de contenido matemático*, 21. 5
- [3] Argüedas, Y., Santana-Cibrian, M., and Velasco-Hernández, J. (2019). Transmission dynamics of acute respiratory diseases in a population structured by age. *Mathematical Biosciences and Engineering*, 16(6):7477–7493. 5, 5
- [4] Bagal, D. K., Rath, A., Barua, A., and Patnaik, D. (2020). Estimating the parameters of susceptible-infected-recovered model of covid-19 cases in india during lockdown periods. *Chaos, Solitons & Fractals*, 140:110154. 1
- [5] Batistela, C. M., Correa, D. P., Buenoc, Á. M., and Piqueira, J. R. (2020). Compartmental model with loss of immunity: analysis and parameters estimation for covid-19. *arXiv preprint arXiv:2007.01295*. 1
- [6] Bettencourt, L. and Ribeiro, R. (2008). Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases. *PlosOne*, 3(5):e2185. 5

- [7] Bliznashki, S. (2020). A Bayesian Logistic Growth Model for the Spread of COVID-19 in New York. *medRxiv*, 14(12). 5
- [8] Boersch-Supan, P., Ryan, S., and Johnson, L. (2017). deBInfer: Bayesian inference for dynamical models of biological systems in R. *Methods in Ecology and Evolution*, 8:511–518. 5
- [9] Brown, G., Porter, A., Oleson, J., and Hinman, J. (2018). Approximate Bayesian computation for spatial SEIR(S) epidemic models. *Spatial and Spatio-temporal Epidemiology*, 24(10):2685–2697. 5
- [10] Calvetti, D., Hoover, A., Rose, J., and Somersalo, E. (2020). Metapopulation network models for understanding, predicting and managing the coronavirus disease covid-19. *arXiv preprint arXiv:2005.06137*. 1
- [11] Capistrán, M., Capella, A., and Christen, A. (2020). Forecasting hospital demand during COVID-19 pandemic outbreaks. *arXiv e-prints*. 5, 5
- [12] Capistrán, M., Moreles, M., and Lara, B. (2009). Parameter Estimation of Some Epidemic Models. The Case of Recurrent Epidemics Caused by Respiratory Syncytial Virus. *Bulletin of Mathematical Biology*, 71(8):1890–1901. 5
- [13] Carpenter, B., Gelman, A., Hoffman, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32. 4, 5, 5, 5, 6, 7, 8, 10
- [14] Chatzilena, A., Leeuwen, E., Ratmann, O., Baguelin, M., and Demiris, N. (2019). Contemporary statistical inference for infectious disease models using Stan. *Epidemics*, 29. 5
- [15] Chávez Hernández, M. V., Juárez Valencia, L. H., and Ríos Solís, Y. A. (2019). Penalization and augmented Lagrangian for o-d demand matrix estimation from transit segment counts. *Transportmetrica A: Transport Science*, 15(2):915–943. 2
- [16] Chen, T.-M., Rui, J., Wang, Q.-P., Zhao, Z.-Y., Cui, J.-A., and Yin, L. (2020). A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infectious Diseases of Poverty*, 9(1):1–8. 1
- [17] Choi, S. and Ki, M. (2020). Estimating the reproductive number and the outbreak size of novel coronavirus disease (COVID-19) using mathematical model in republic of korea. *Epidemiology and Health*, pages 123–145. 1
- [18] Chowell, G. (2017). Fitting dynamic models to epidemic outbreak with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2:379–398. 5, 5
- [19] Christen, J. and Fox, C. (2010). A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Anal.*, 5:263–282. 5, 11
- [20] Coletti, P., Libin, P., Petrof, O., Willem, L., Steven, A., Herzog, S. A., Faes, C., Wambua, J., Kuylen, E. J., Beutels, P., et al. (2020). A data-driven metapopulation model for the belgian covid-19 epidemic: assessing the impact of lockdown and exit strategies. *medRxiv*. 1

- [21] Coly, S., Yao, A., Abrial, D., and Garrido, M. (2016). Distributions to model overdispersed count data. *Journal de la Societe Francaise de Statistique*, 157(2):39–63. 5
- [22] Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*:1695. 2.1
- [23] de Anda-Juregui, G. (2020). COVID-19 in Mexico: A Network of Epidemics. *arXiv e-prints*. 2.1
- [24] de la Ciudad de México, G. (2020). Afluencia preliminar en transporte publico. <https://datos.cdmx.gob.mx/explore/dataset/afluencia-preliminar-en-transporte-publico/table/>. 1, 2
- [25] de Mexico, G. (2020). Covid-19 Mexico. <https://coronavirus.gob.mx/datos/>. 1, 5
- [26] Diekmann, O., Heesterbeek, J. A. P., and Metz, J. A. (1990). 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*, 28(4):365–382. 4
- [27] Google (2020). Informes de movilidad local sobre el covid-19. <https://www.google.com/covid19/mobility/>. 1, 2
- [28] Grinsztajn, L., Semenova, E., Margossian, C., and Riou, J. (2020). Bayesian workflow for disease transmission modeling in Stan. *arXiv e-prints*. 5, 5
- [29] House, T., Ford, A., Lan, S., Bilson, S., Buckingham-Jeffery, E., and Girolami, M. (2016). Bayesian uncertainty quantification for transmissibility of influenza, norovirus and Ebola using information geometry. *J.R.Soc. Interface*, 13. 5
- [30] INEGI (2017). Encuesta origen destino en hogares de la zona metropolitana del valle de mexico (eod) 2017. <https://www.inegi.org.mx/programas/eod/2017/>. 1, 2, 2
- [31] INEGI (2020). Instituto nacional de estadística y geografía. <https://www.inegi.org.mx/temas/estructura/default.html#Publicaciones/>. 2
- [32] Izquierdo, L. D. et al. (2020). *Informe técnico nuevo coronavirus 2019-nCoV*. PhD thesis, Instituto de Salud Carlos III. 1
- [33] Keeling, M. and Rohani, P. (2002). Estimating spatial coupling in epidemiological systems: a mechanistic approach. *Ecology Letters*, 5:20–29. 1
- [34] Keeling, M. J. and Rohani, P. (2011). *Modeling infectious diseases in humans and animals*. Princeton University Press. 1
- [35] Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., and Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science*, 368(6490):489–493. 1, 3, 3
- [36] Liu, F., Li, X., and Zhu, G. (2020). Using the contact network model and metropolis-hastings sampling to reconstruct the covid-19 spread on the diamond princess. *Science Bulletin*. 1

- [37] Luzyanina, T. and Bocharov, G. (2018). Markov chain Monte Carlo parameter estimation of the ODE compartmental cell growth model. *Mathematical Biology and Bioinformatics*, 13:376–391. [5](#)
- [38] Ma, Z. (2020). Spatiotemporal fluctuation scaling law and metapopulation modeling of the novel coronavirus (covid-19) and sars outbreaks. *arXiv preprint arXiv:2003.03714*. [1](#)
- [39] Magal, P. and Webb, G. (2018). The Parameter Identification Problem for SIR Epidemic Models: Identifying Unreported Cases. *Journal of Mathematical Biology*, 77:1629–1648. [5](#)
- [40] McBryde, E., Gibson, G., Pettitt, A., Zhang, Y., Zhao, B., and McElwain, D. (2006). Bayesian modelling of an epidemic of severe acute respiratory syndrome. *Bulletin of mathematical biology*, 68(4):889–917. [1](#)
- [41] Miller, A., Foti, N., Lewnard, J., Jewell, N., Guestrin, C., and Fox, E. (2020). Mobility trends provide a leading indicator of changes in SARS-CoV-2 transmission. *medRxiv*. [1](#)
- [42] Nayens, T. and Faes, C. and Molenberghs, G. (2012). A generalized Poisson-gamma model for spatially overdispersed data. *Spatial and Spatio-temporal Epidemiology*, 3:185–194. [5](#)
- [43] Nesteruk, I. (2020). Statistics-based predictions of coronavirus epidemic spreading in mainland china. *medRxiv*. [1](#)
- [44] Nour, M., Cömert, Z., and Polat, K. (2020). A novel medical diagnosis model for covid-19 infection detection based on deep features and bayesian optimization. *Applied Soft Computing*, page 106580. [1](#)
- [45] Paules, C. I., Marston, H. D., and Fauci, A. S. (2020). Coronavirus infections more than just the common cold. *Jama*, 323(8):707–708. [1](#)
- [46] Pei, S. and Shaman, J. (2020). Initial Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US. *medRxiv*. [1](#)
- [47] Prieto, K. (2020). Current forecast of covid-19: a bayesian and machine learning approaches. *medRxiv*. [1](#), [5](#)
- [48] Prieto, K. and Dorn, O. (2016). Sparsity and level set regularization for diffuse optical tomography using a transport model in 2D. *Inverse Problems*, 33(1):014001. [5](#)
- [49] Prieto, K. and Ibarguen-Mondragon, E. (2019). Parameter estimation, sensitivity and control strategies analysis in the spread of influenza in Mexico. *Journal of Physics: Conference Series*, 1408(1):012020. [5](#)
- [50] Roosa, K. and Chowell, G. (2019). Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. *Theoretical Biology and Medical Modelling*, 16(1). [5](#)
- [51] Savi, P. V., Savi, M. A., and Borges, B. (2020). A mathematical description of the dynamics of coronavirus disease (COVID-10): A case study of Brazil. *arXiv preprint arXiv:2004.03495*. [1](#)

- [52] Şenel, K., Özding, M., Öztürkcan, D. S., and Akgül, A. (2020). 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*, 40(2):127–131. [1](#)
- [53] Shaikh, A. S., Shaikh, I. N., and Nisar, K. S. (2020). A mathematical model of COVID-19 using fractional derivative: Outbreak in india with dynamics of transmission and control. *Preprints*. [1](#)
- [54] Smirnova, A., DeCamp, L., and Liu, H. (2016). *Inverse Problems and Ebola Virus Disease Using an Age of Infection Model*, pages 103–121. Springer, Cham. [5](#)
- [55] Spiess, H. (1987). A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B: Methodological*, 21(5):395–412. [2](#)
- [56] Stojanović, O., Leugering, J., Pipa, G., Ghazzi, S., and Ullrich, A. (2019). A Bayesian Monte Carlo approach for predicting the spread of infectious diseases. *PLoS ONE*, 14(12). [5](#)
- [57] Wells, K. and Lurgi, M. (2020). Covid-19 containment policies through time may cost more lives at metapopulation level. *medRxiv*. [1](#)
- [58] World Health Organization (2020). Novel coronavirus (2019-ncov) situation reports. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation>. Acces 05-10-2020. [1](#)
- [59] Worldometer (2020). Worldometer Coronavirus Updates. <https://www.worldometers.info/>. Acces 05-10-2020. [2](#)
- [60] Yang, C. and Wang, J. (2020). A mathematical model for the novel coronavirus epidemic in wuhan, china. *Mathematical Biosciences and Engineering*, 17(3):2708–2724. [1](#)
- [61] Yang, W., Zhang, D., Peng, L., Zhuge, C., and Hong, L. (2020). Rational evaluation of various epidemic models based on the covid-19 data of china. *arXiv preprint arXiv:2003.05666*. [1](#)
- [62] Zhao, Z., Zhu, Y.-Z., Xu, J.-W., Hu, Q.-Q., Lei, Z., Rui, J., Liu, X., Wang, Y., Luo, L., Yu, S.-S., et al. (2020). A mathematical model for estimating the age-specific transmissibility of a novel coronavirus. *medRxiv*. [1](#)
- [63] Zhong, L., Mu, L., Li, J., Wang, J., Yin, Z., and Liu, D. (2020). Early prediction of the 2019 novel coronavirus outbreak in the mainland china based on simple mathematical model. *Ieee Access*, 8:51761–51769. [1](#)
- [64] Zhou, C. (2020). 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*. [1](#)