Study of Basic Reproduction Number Projection of SARS-CoV-2 Epidemic in USA and Brazil

Erika Severeyn
Dept. of Thermodynamics and Transfer Phenomena
Simón BolívarUniversity
Caracas, Venezuela
severeynerika@usb.ve
0000-0002-9500-3532

Sara Wong
Dept. of Electronics and Circuits
Simón BolívarUniversity
Caracas, Venezuela
swong@usb.ve
0000-0002-8999-2653

Héctor Herrera
Dept. of Biological and Biochemical Process Technology
Simón BolívarUniversity
Caracas, Venezuela
haherrera@usb.ve
0000-0003-0278-3427

Alexandra La Cruz
Faculty of Engineering
Ibagué University
Ibagué, Colombia
alexandra.lacruz@unibague.edu.co
0000-0001-6052-2933

Jesús Velásquez
Dept. of Thermodynamics and Transfer Phenomena
Simón BolívarUniversity
Caracas, Venezuela
jmvelasquezf@usb.ve
0000-0002-0811-3320

Mónica Huerta
Carrera de Telecomunicaciones
Universidad Politécnica Salesiana
Cuenca, Ecuador
mhuerta@ups.edu.ec
0000-0003-4435-7987

***Abstract*—In December 2019, a group of patients presented a diagnosis of pneumonia of unknown etiology in Hubei Province, Wuhan, China. By January 2020, authorities around the world faced a new coronavirus (SARS-CoV-2). By August 2020, the two countries with the highest number of SARS-CoV-2 infections are the USA and Brazil. The transmission rate of a virus is studied from the basic reproduction number (R_0). The SIR model is the simplest compartmental epidemiological model (Susceptible, Infectious and Recovered). The SIR model can be used to estimate R_0 by fitting the curve of the infected compartment to the experimental curve of infected subjects per day. The aim of this work is to study the projection of the R_0 of SARS-CoV-2 in the USA and Brazil. For this purpose, five experiments were performed by adjusting the SIR model curve of infected compartment to experimental data at five time intervals (the first 14, 28, 42, 56 and 187 days for the USA data, and 177 days for Brazil data). In the first two time intervals the R_0 varied between 5.46 and 7.75 for the USA data and 1.84 and 4.29 for Brazil data, and in the last three time intervals the R_0 decreased to 1.05 for the USA data and 1.01 for Brazil data, suggesting that the social distancing measures implemented in both countries were able to decrease the infection spreading. The differences in the R_0 values of the five experiments imply that R_0 also depends on the preventive measures implemented to face the pandemic.**

Keywords— SARS-CoV-2, Mathematical Epidemiologic Model, Parametric Fitting.

I. INTRODUCTION

In December 2019, a group of patients presented to different hospitals with diagnoses of pneumonia of unknown etiology in Hubei Province, Wuhan, China [1, 2]. Most of these patients were epidemiologically associated to a wholesale market for live and unprocessed fish, shellfish and live animals in Hubei province [3, 4].

The economic growth of China has led to a high demand for animal protein, including exotic animals, such as snakes, civets, pangolins and bats [5]. Deficiencies in bio-security measures in food markets have allowed viruses to be transmitted between animals and from animals to humans (zoonosis) [6]. An initial lack of knowledge regarding the management of infected patients and international air traffic had contributed to the accelerated global spread of SARS in 2002-2003 [7].

Between December 18 and 29, 2019, the first five cases were reported; four of these patients were hospitalized for presenting acute respiratory distress syndrome and one of them died. Most patients reported a direct or indirect association with a food market in Hubei Province in Wuhan [8]. Because of the closure of the Wuhan market in January 1, 2020, there was no clear evidence of person-to-person transmission. By January 2, a total of 41 patients had been hospitalized and only one patient with serious pre-existing conditions had died. In January 7, Chinese authorities announced that they had identified a new type of coronavirus (SARS-CoV-2) [9]. Simultaneously, other possible pathogens were discarded, including the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) [7], the Middle East Respiratory Syndrome coronavirus (MERS-CoV) [10], the influenza virus [11], the avian influenza virus [12] and the adenoviruses [13]. From this point on, the authorities worldwide knew that they faced a new threat [14].

When a new virus appears, the first characteristic to look for is the way it is transmitted. The transmission rate of a virus is studied from the basic reproduction number (R_0) [15]. Once this information is available, projections of how the virus will be transmitted in a population can be made. Models that simulate the transmission of infectious diseases are used; among these models is the simplest one called SIR model. The SIR model is the simplest compartmental model; it is based in three compartments (Susceptible, Infectious and Recovered) [16]. In the Susceptible (S) compartment are subjects who may be infected, in the Infectious (I) compartment are subjects who have the infection and in the Recovered (R) compartment are subjects who have dead or recovered from the infection (now have immunity)[16].

Initial estimates from Chinese data in January suggested that the number of people infected doubled every 6 or 7 days, with an R_0 of 2.24 to 3.58 [17]. In contrast, data presented in early April suggested that SARS-CoV-2 is spreading faster and easier than originally anticipated. Results published by the epidemiological team at Los Alamos National Laboratory show an average R_0 of 5.7 [18]. Two different strains have

also been reported since the appearance of the disease, one of which spread in Italy, Spain and the United States and which is speculated to have a greater ability to spread [5, 19].

Currently, the two countries with the highest number of SARS-CoV-2 infections are the United States of America (USA) and Brazil, with 5,701,878 [20] and 3,460,413 [21] cases of infected by August 20 2020, respectively. The first known case of SARS-CoV-2 in the USA was confirmed on January 21, 2020, in a 35-year-old returnee from Wuhan, China [22]. Two days later, restrictions on travelers arriving from China were announced. The White House advised against any gathering of more than 10 people on March 13, and a policy of social distancing has since been established [22]. In February 25, the Brazilian Ministry of Health confirmed the first case in Brazil and South American region, and the State of São Paulo declared a state-wide quarantine since March 24 [23].

One of the methods for estimating R_0 consists of fitting the curve generated by the SIR model of infected subjects per day to the actual curve of infected subjects per day, as reported in [20, 21]. The projection of R_0 will depend largely on the data, therefore, when the infection begins, it only has the data from the first days and the estimation of R_0 will be based on those data, once governments start taking preventive measures the infected curve will be modified and therefore the R_0 is different.

The aim of this work is to study the different projections of the R_0 value of the SARS-CoV-2 in the USA and Brazil. For this purpose, the SIR has been fitted to different time periods of the existing data on infected cases. In the following section the methodology will be explained. Results and discussion will be presented in Sections III and IV. Additionally, conclusions and proposals for future work will be presented in Section V.

II. Methodology

A. Database

The database was collected from [20, 21], the data used correspond to the number of cases of SARS-CoV-2 infection per day. Currently, the USA and Brazil have the highest number of infected in the world. This database started reporting active cases since February 15; however the first case of SARS-CoV-2 were reported in the USA in January 21. It suggests that the amount of infected subjects by February 15 correspond to all the cases reported between January 21 and February 15. In the case of Brazil the database cover from the day the first case was reported in February 25.

B. SIR Model

The SIR (Susceptible-Infected-Recovered) model, developed by Ronald Ross, William Hamer, and others in the early 20th century, consists of a system of three coupled nonlinear ordinary differential equations, which does not possess an explicit formula solution [24]. The SIR model corresponds to the follow three equations:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \quad (1)$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$

Where:

- β is the infection rate (the rate of the population susceptible become infected).
- γ is the removal rate (the rate of infected population recovering or dying), with the inverse of γ can be calculate the recovery period of an infected subject since the moment infection begins until the negative diagnostic test result.
- S is susceptible population per day.
- I is the infected population per day.
- R is the recovered population per day.

The basic reproduction number is a theoretical parameter that provides some information about the speed with which a disease can spread in a given population. The R_0 can be calculated from the relation between β and γ in the SIR model [25]:

$$R_0 = \frac{\beta}{\gamma} \quad (4)$$

The dynamics of the infectious group depends on R_0, if $R_0 \geq 1$, the infection will spread, if $R_0 < 1$ the infection will be extinguish.

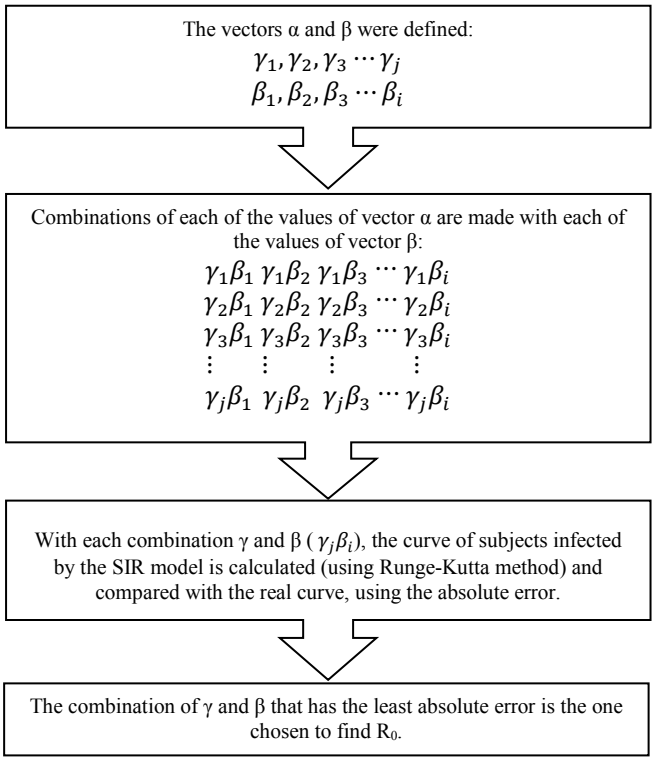

Fig. 1. Fitting of γ and β parameters of the SIR model to infected experimental curve.

C. Parameter fitting

The fitting of γ and β parameters of the SIR model to the experimental data was performed following the procedure showed in Figure 1 [26]. The vectors of γ and β were set as inputs to the algorithm. At first, an attempt was made to limit

these vectors, taking as data the R_0 (between 2 and 3.1) [27] and the recovery time (between 8 to 21 days) [28]. However, these limitations prevented to achieve the minimization of the objective function, therefore very high and very low values of gamma and beta were established as limits and their values were plotted as a function of absolute error to verify the minimum absolute error. Vectors were constructed with an equidistance of 10^{-4}.

Combinations were made with each of the values of the vectors γ and β. With each combination, the curve of infected subjects was found by solving the differential equation system of the SIR model using a six-stage of the fifth-order Runge-Kutta method, and this curve was compared with the experimental curve of infected subjects found in [20, 21] using the absolute error as an objective function. Finally, the combination of γ and β that has the lowest absolute error is the one chosen to find R_0.

In this study, five experiments were performed to look for the projection of the susceptible, infected and recovered curves of the five time intervals:

- Experiment 1: Active cases per day from the first day (February 15 for the USA and February 25 for Brazil) to fourteenth day were used for the projection
- Experiment 2: Active cases per day from the first day (February 15 for the USA and February 25 for Brazil) to twenty eighth day were used for the projection.
- Experiment 3: Active cases per day from the first day (February 15 for the USA and February 25 for Brazil) to forty second day were used for the projection.
- Experiment 4: Active cases per day from the first day (February 15 for the USA and February 25 for Brazil) to fifty sixth day were used for the projection.
- Experiment 5: Active cases per day from the first day (February 15 for the USA and February 25 for Brazil) to one hundred and eighty seventh day for the USA and one hundred and seventy seventh day for Brazil were used for the projection.

In each experiment, the procedure explained in Figure 1 was performed for the curve of infected cases in Brazil and the USA.

III. Results

Table I shows the projections for R_0, β, γ and recovery period of the five experiments for the USA and Brazil. In each case, the infected curve was fitted using the number of days of the experimental curve, for example for the first experiment; the experimental curve was used from day one to day 14. This was done in this way to observe how R_0 projection changes at different time points.

Figures 2a, 2b, 2c, 2d and 2e show the graphs of the fit of the SIR model curves for the USA data and in Figures 3a, 3b, 3c, 3d and 3e the SIR model curves for Brazil data. In the SIR model curve fit graphs, the susceptible, infected and recovered curves are shown on the left and the infected curve is zoomed in on the right for more detail.

IV. Discussion

In Table I it can be seen that in experiments 1, 2 and 3, the R_0 fluctuates between 4.57 and 7.73 for the USA data and between 1.84 and 4.29 for Brazil data, this tendency changes in experiments 4 and 5 where the R_0 fell to 1.05 for the USA data and 1.01 for the Brazilian data, these findings suggest that social distancing, stipulated on March 13 for the USA government and on March 24 for the Brazil government, had a significant impact on the decrease of R_0.

The R_0 found in this research for the USA and Brazil is higher than the reported by China at the beginning of the epidemic, this could be due to the existence of at least two strain of SARS-CoV-2 and the strain that developed in the USA, Italy, Spain and Brazil, is more contagious than the strain that developed in China [5]. In any case, it should be taken as a starting point that SARS-CoV-2 has an R_0 between 4.29 and 7.73 for the purposes of future projections and preventive health measures.

A directly proportional relationship was observed between R_0 and the projection of the maximum number of infected subjects in the USA and Brazilian data. In Figures 2a, 2b and 2c the maximum number of active infected cases in the USA ranged from 1.5×10^8 to 2×10^8 subjects, after the establishment of social distancing the maximum number decreased and ranged from 4×10^5 to 2×10^6 subjects. In Figures 3a, 3b and 3c, the maximum number of active cases of infection in Brazil ranged from 2.5×10^7 to 10^8 subjects, after the establishment of social distancing the maximum number decreased and ranged from 2×10^4 to 8.5×10^5 subjects. All these findings suggest that social distancing could prevent millions of deaths.

On the other hand, it can be seen that the duration of the epidemic also varies according to R_0, the duration in Figures 2a, 2b and 2c ranged from 200 to 400 days, while in Figures 2d and 2e ranged from 100 and 250 days, the duration in figures 3a, 3b and 3c ranged from 100 and 200 days, while in figures 3d and 3e ranged from 80 and 200 days, if the applied social distancing conditions implemented are maintained in the USA and Brazil, they could emerge from the pandemic within approximately 63 days and 73 days, respectively.

The model has certain limitations, it can be observed that the sum of absolute error increases as the number of days studied increases in both data sets (Brazil and USA), this may be due to the fact that when there are more days the error of the additional day is added to the sum of the absolute error values. Also, the sum of the absolute errors in the experiment 2, 3 and 4 increased modestly, which changed with the experiment 5 that the error increased by 75-fold for the USA data and 19-fold for Brazil data, compared to the experiment 4. This could be due to the fact that the model being used does not take into account certain aspects such as the amount of population varies every day by the number of people who naturally (not connected to SARS-CoV-2) died and were born, and there is also a percentage of those recovered (14%) who could be re-infected by the disease [29]. In South-Korea 91 cases of SARS-CoV-2 recovery were actively re-infected, this type of population should be returned to the group of susceptible, and this is not contemplate by the SIR model used in this research [30].

From the model, the average number of days a patient recovers can be calculated, as explained in methodology section. Table I shows that the recovery period ranged from 38.462 to 38.536 days in experiments 1 and 2 for the USA data-set, and ranged from 3.125 and 14.286 for Brazil data set. On the other hand, in experiments 3, 4 and 5 the recovery period ranged from 0.252 and 16 days for the USA data set, and 0.064 and 6.524 days for the Brazilian data set. This

TABLE I. R_0, γ AND β PROJECTIONS TO EACH EXPERIMENT.

Country	Parameters	Experiment				
		1	*2*	*3*	*4*	*5*
USA	R_0	5.399	7.737	4.575	1.054	1.143
	γ	0.026	0.026	0.060	3.966	0.560
	β	0.142	0.201	0.275	4.180	0.640
	Recovery period [days]	38.023	38.536	16.611	0.252	1.786
	Sum of absolute error	5.492	101.360	1398.263	5075.930	508164.234
Brazil	R_0	4.286	1.844	2.499	1.013	1.097
	γ	0.070	0.320	0.153	15.700	1.030
	β	0.300	0.590	0.383	15.900	1.130
	Recovery period [days]	14.286	3.125	6.524	0.064	0.971
	Sum of absolute error	2.105	20.572	647.436	2745.602	135610.144

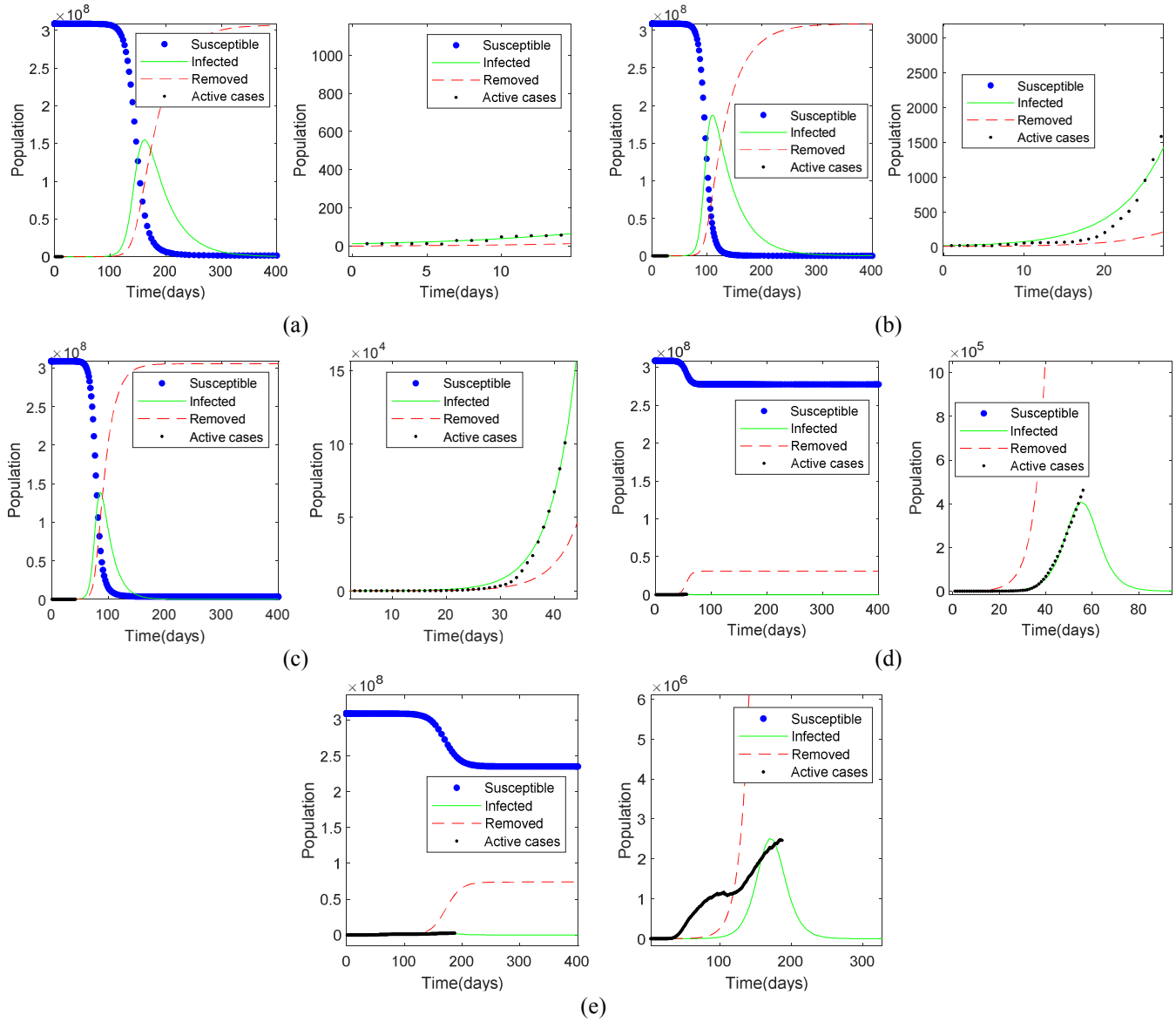

Fig. 2. USA SIR model simulation and zoom in on the experimental data interval: (a) Experiment 1, (b) experiment 2, (c) experiment 3, (d) experiment 4 and (e) experiment 5.

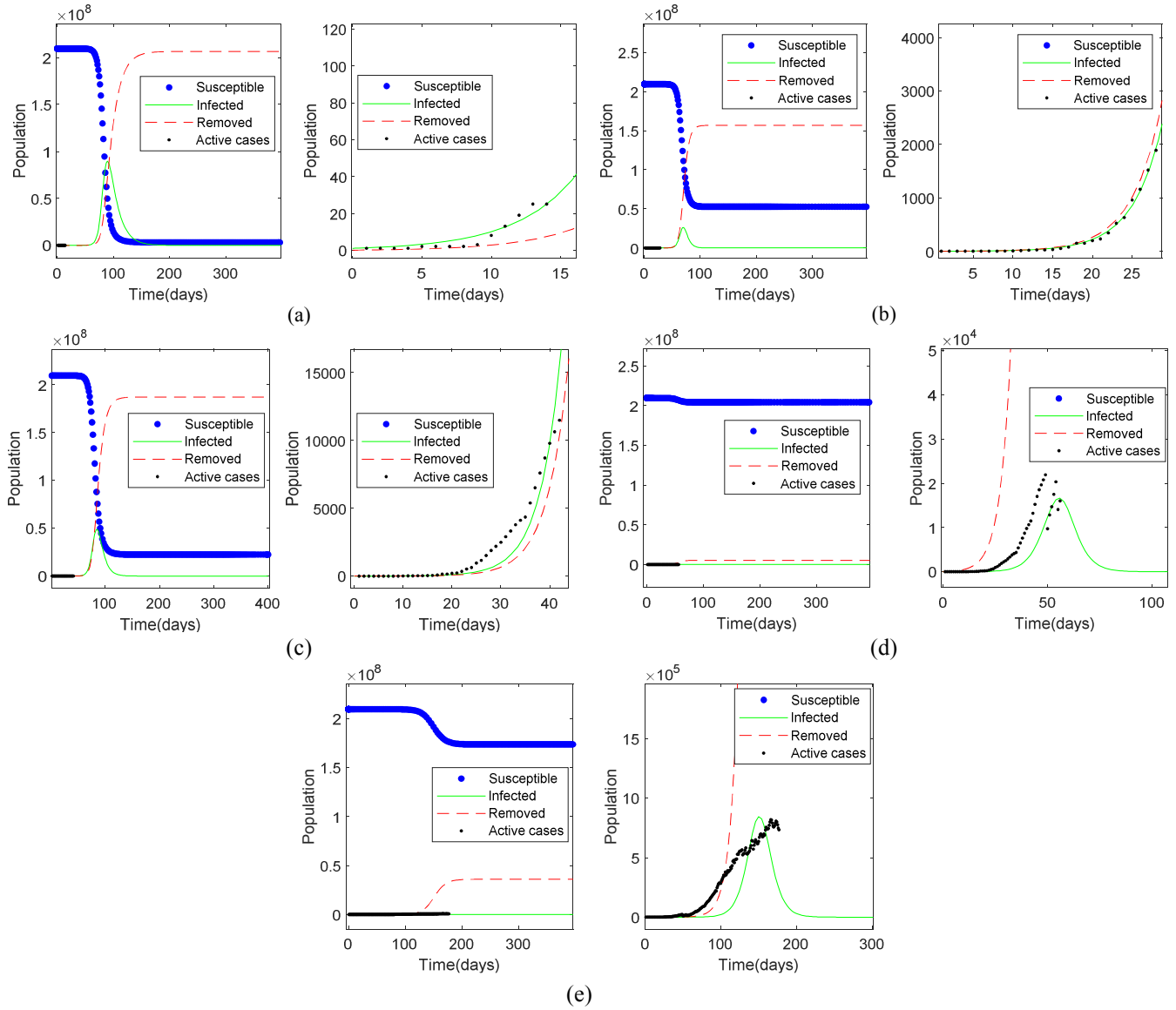

Fig. 3. Brazil SIR model simulation and zoom in on the experimental data interval: (a) Experiment 1, (b) experiment 2, (c) experiment 3, (d) experiment 4 and (e) experiment 5.

could be due to that the probability of infection transmission decreased as a consequence of establishing social distancing, and the natural values of the virus metabolism are modified by the new conditions, therefore, only the values prior to the establishment of social distancing (experiments 1 and 2 for the Brazilian and USA datasets) should be taken to understand the dynamics of natural SARS-CoV-2 transmission. It can also be verified that the recovery period found in the USA dataset ($\approx$38 days) is higher than that found in Brazil and the literature [27] (between 8 and 21 days) suggests an increased recovery period when the amount of transmission is higher.

V. Conclusions

In this study the variations in R_0 were assessed, using five time intervals as experimental values for the fitting of the infected curve. The decrease in R_0 was observed after the establishment of social distancing, suggesting that the distancing measures implemented in the USA and Brazil were able to decrease the ability of the infection spread.

The nature of SARS-CoV-2 gives it an immense ability to spread because it can be transmitted by saliva and nasal fluids and can survive on surfaces for days. This makes distance measurements very important in flattening the infected curve and preventing the collapse of the health care system.

The differences in R_0 values for the five experiments suggest that R_0 depends not only on the nature of the infection but also on the economic activity of the country and the preventive measures taken against the pandemic. Therefore, the R_0 could be different in each country. For future works, R_0 values will be calculated for other countries with different economic developments and with different cultures.

Acknowledgments

This work was funded by the Research and Development Deanery of Salesian Polytechnic University and the Research and Development Deanery of the Simón Bolívar University (DID).

REFERENCES

[1] C. Wang, P. W. Horby, F. G. Hayden and G. F. Gao. "A novel coronavirus outbreak of global health concern". *The Lancet*, vol. 395, no. 10223, pp. 470-473, 2020.

[2] I. Bogoch, A. Watts, A. Thomas-Bachli, C. Huber, M.U.G. Kraemer and K. Khan. "Pneumonia of unknown etiology in wuhan, China: potential for international spread via commercial air travel". *J. Trav. Med*, vol. 2, no.1, pp. 1-3, 2020.

[3] H. Lu, C. W. Stratton and Y. W. Tang. "Outbreak of Pneumonia of Unknown Etiology in Wuhan China: the Mystery and the Miracle". *Journal of Medical Virology*, vol. 92, no. 4, pp. 401-402, 2020.

[4] V. C. Cheng, S. K. Lau, P. C. Woo and K. Y. Yuen. "Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection". *Clinical microbiology reviews*, vol. 20, no. 4, pp. 660-694, 2007.

[5] K. G. Andersen, A. Rambaut, W. I.Lipkin, E. C. Holmes and R. F. Garry. "The proximal origin of SARS-CoV-2". *Nature medicine*, vol. 26, no.4, pp. 450-452, 2020.

[6] A. Du Toit. "Outbreak of a novel coronavirus". *Nature Reviews Microbiology*, vol. 18, no. 3, pp. 123-124, 2020.

[7] S. K. Lau, P. C. Woo, K. S. Li, Y. Huang, H. W. Tsoi, B. H. Wong, and K. Y. Yuen. "Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats". *Proceedings of the National Academy of Sciences*, vol. 102, no. 39, pp. 14040-14045, 2005.

[8] L. L. Ren, Y. M. Wang, Z. Q. Wu, Z. C. Xiang, L. Guo, T. Xu and H. Li. "Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study". *Chinese medical journal*, vol 133, no. 9, pp. 1015.1024, 2020.

[9] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu and Z. Cheng. "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China". *The lancet*, vol. 395, no.10223, pp. 497-506, 2020.

[10] R. J. de Groot, S. C. Baker, R. S. Baric, C. S. Brown, C. Drosten, L. Enjuanes and S. Perlman. "Commentary: Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group". *Journal of virology*, vol. 87, no. 14, pp. 7790-7792, 2013.

[11] M. Lipsitch, L. Finelli, R. T. Heffernan, G. M. Leung and Redd; for the 2009 H1N1 Surveillance Group, S. C. "Improving the evidence base for decision making during a pandemic: the example of 2009 influenza A/H1N1". Biosecurity and bioterrorism: biodefense strategy, practice, and science, vol. 9, no. 2, pp. 89-115, 2011.

[12] E. C. Claas, A. D. Osterhaus, R. Van Beek, J. C. De Jong, G. F. Rimmelzwaan, D. A. Senne and R. G. Webster. "Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus". *The Lancet*, vol. 351, no. 9101, pp. 472-477, 1998.

[13] N. Tatsis and H. C. Ertl. "Adenoviruses as vaccine vectors. *Molecular Therapy"*, vol. 10, no. 4, pp. 616-629, 2004.

[14] J. H. Tanne, E. Hayasaki, M. Zastrow, P. Pulla, P. Smith and A. G. Rada. "Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide", BMJ, vol. 368, 2020.

[15] F. M. Guerra, S. Bolotin, G. Lim, J. Heffernan, S. L. Deeks, Y. Li and N. S. Crowcroft. "The basic reproduction number (R_0) of measles: a systematic review". *The Lancet Infectious Diseases*, vol. 17, no. 12, pp. e420-e428, 2017.

[16] H. H. Weiss. "The SIR model and the foundations of public health". *Materials matematics*, vol. 2013, no. 3, pp.1-17 (2013).

[17] S. Zhao, Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang and M. H. Wang. "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak". *International journal of infectious diseases*, vol. 92, pp. 214-217, 2020.

[18] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner and R. Ke. "High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2". E*merging infectious diseases*, vol. 26, no. 7, pp.1-10, 2020.

[19] I. M. Ibrahim, D. H. Abdelmalek, M. E. Elshahat and A. A. Elfiky. "COVID-19 spike-host cell receptor GRP78 binding site prediction". *Journal of Infection*, vol 80, no. 5, pp. 554-562, 2020.

[20] Worldometers Home page. https://www.worldometers.info/coronavirus/country/us/, last accessed 2020/08/20.

[21] Worldometers Home page. https://www.worldometers.info/coronavirus/country/brazil/, last accessed 2020/08/20.

[22] M. L. Holshue, C. DeBolt, S. Lindquist, K. H. Lofy, J. Wiesman, H. Bruce and G. Diaz. "First case of 2019 novel coronavirus in the United States". *New England Journal of Medicine*, vol 382, pp. 929-936, 2020.

[23] A. J. Rodriguez-Morales, V. Gallego, J. P. Escalera-Antezana, C. A. Méndez, L. I. Zambrano, C. Franco-Paredes and A. Risquez. "COVID-19 in Latin America: The implications of the first confirmed case in Brazil". *Travel medicine and infectious disease*, vol. 101613, pp. 1-3, 2020.

[24] R.M. Anderson. "Discussion: the Kermack-McKendrick epidemic threshold theorem". *Bulletin of mathematical biology*, vol. 53, no. 1,pp. 1-32, 1991.

[25] G. G. Alcaraz and C. Vargas-De-León. "Modeling control strategies for influenza A H1N1 epidemics: SIR models". *Revista Mexicana de Física*, vol. 58, no. 1, pp. 37-43, 2012.

[26] G. Schneckenreither, N. Popper, G. Zauner and F. Breitenecker. "Modelling SIR-type epidemics by ODEs, PDEs, difference equations and cellular automata–A comparative study". *Simulation Modelling Practice and Theory*, vol.16, no. 8, pp. 1014-1023, 2008.

[27] Y. Liu, A. A. Gayle, A. Wilder-Smith and J. Rocklöv. "The reproductive number of COVID-19 is higher compared to SARS coronavirus". *Journal of travel medicine*, vol 27, no.2, pp.1-4, 2020.

[28] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, and A. Dighe. "Estimates of the severity of coronavirus disease 2019: a model-based analysis". *The Lancet Infectious Diseases*, vol. 20, no.6, pp. 669-677, 2020.

[29] L. Lan, D. Xu, G. Ye, C. Xia, S. Wang, Y. Li and H. Xu. "Positive RT-PCR test results in patients recovered from COVID-19". *JAMA*, vol. 323, no. 15, pp. 1502-1503, 2020.

[30] E. Mahase. "Covid-19: WHO and South Korea investigate reconfirmed cases". *BMJ*, vol. m1498, pp. 1-2, 2020.