COVID-19 Trend Analysis in Mexican States and Cities

Henrique Mohallem Paiva[1], *Senior Member, IEEE*, Rubens Junqueira Magalhães Afonso[2], *Member, IEEE*,
Davi Gonçalves Sanches[1], and Frederico José Ribeiro Pelogia[1]

hmpaiva@unifesp.br, rubensjm@ita.br, davi.sanches@unifesp.br, fpelogia@unifesp.br

[1] ICT – Instituto de Ciência e Tecnologia (*Institute of Science and Technology*)
UNIFESP – Universidade Federal de São Paulo (*Federal University of Sao Paulo*)
Rua Talim, 330, São José dos Campos, SP, Brazil, 12231-280

[2] Divisão de Engenharia Eletrônica (*Electronics Engineering Division*)
ITA – Instituto Tecnológico de Aeronáutica (*Aeronautical Institute of Technology*)
Praça Marechal Eduardo Gomes, 50, São José dos Campos, SP, Brazil, 12228-900

Abstract—This paper presents a trend analysis of the COVID-19 pandemics in Mexico. The studies were run in a subnational basis because they are more useful that way, providing important information about the pandemic to local authorities. Unlike classic approaches in the literature, the trend analysis presented here is not based on the variations in the number of infections along time, but rather on the predicted value of the final number of infections, which is updated every day employing new data. Results for four states and four cities, selected among the most populated in Mexico, are presented. The model was able to suitably fit the local data for the selected regions under evaluation. Moreover, the trend analysis enabled one to assess the accuracy of the forecasts.

Index Terms—epidemiology, mathematical model, trend analysis, COVID-19, SARS-CoV-2

I. Introduction

The SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) is the new coronavirus that causes the COVID-19 infection. Primarily understood as a lung-tropic virus that infects the respiratory tract, more research on its physiopathology has shown that the COVID-19 infection presents itself as a general inflammatory disease in humans, since its receptor ACE2 is present in the cells membrane of many different human tissues, such as the gastrointestinal epithelial cells [1], platelets and endothelial cells [2], neurons [3], among others. For this reason, COVID-19 symptoms are not only respiratory, but can include diarrhea, nausea, neurological complications and atherothrombosis [1]–[3].

Since the first outbreak in Wuhan, China, in December 2019, studies have shown that the most effective measures to prevent the COVID-19 infection are social distancing, using hand sanitizers and wearing personal protective equipment (PPE), such as masks [4]. It is also known that media trust and social norms have a major influence over the adoption of safety behaviors by the population [5].

Mathematical modeling in epidemics is a reasonable approach when studying the disease's dynamics because it can provide accurate information about new cases, new deaths and serve as a tool for authorities to predict the epidemic curve [6]. Models can include data and many different compartments to better estimate the population's parameters and predict the evolution of the epidemic [6]–[8].

Mexico is one of the most affected countries by COVID-19 since it has the third-largest death toll in the world until now. Mexico has the third-largest population in the American continent and faces socioeconomic disparities that influence the impacts of the pandemic [9]. Studies show that cases from the first wave in early 2020 were imported in February and in a few weeks' time community transmission was also reported, with new cases and deaths concentrated in Mexico City [10]. The relaxation of the safety measures after the first wave led to a greater second wave of cases that were no longer concentrated in the country's capital [11]. Mathematical modeling of the pandemic in Mexico can help the country's health authorities to overcome its difficulties as regards testing and predicting new cases and deaths.

In this paper, the final number of deaths is forecast based on local data for selected Mexican states and cities and trend analysis is applied to evaluate the accuracy of the fitted model. For the selected localities, the model has been able to represent suitably the data and the trend analysis provided an adequate figure of merit to assess the validity of the forecast and to support the local authorities in their decision-making process.

II. The Proposed Approach

This paper describes the cumulative number of infections $C(t)$ using Richards growth model, which is characterized as [12], [13]:

$$\frac{dC}{dt}(t) = \frac{1}{\delta\nu}C(t)\left(1 - \frac{(C(t))^{\nu}}{A^{\nu}}\right) \qquad (1)$$

subject to

$$C(t_p) = A\left(\frac{1}{1+\nu}\right)^{\frac{1}{\nu}} \qquad (2)$$

where t is the time, A is the number of infected people at the end of the epidemic (final number of infections), $(\delta\nu)^{-1}$ is the intrinsic growth rate, ν is a parameter associated to the asymmetry between the acceleration and deceleration phases of the pandemic, and t_p is the turning point between these two phases. The parameters A, t_p, ν, δ are strictly positive constant real numbers. C and A represent numbers of infected people, t, t_p and δ are measured in days, and ν is an dimensionless value.

It should be noted that the cumulative number of infections is not a real value, but rather an integer one, and therefore should be represented by a discrete function. Similarly, although the time values are continuous, they are reported as discrete values, for the data update occurs once a day. However, as is costumary in epidemiological models, this discrete behavior is approximated here by a continuous function, without loss of representativity.

The model described by eqs. (1)-(2) has an analytical solution, given by [12]–[14]:

$$C(t) = A\frac{1}{\left(1 + \nu e^{-\frac{(t-t_p)}{\delta}}\right)^{\frac{1}{\nu}}} \tag{3}$$

An advantage of this closed form is that it simplifies theoretical and numerical analyses [12], for it does not require the solution of a differential equation.

By differentiating $C(t)$ in (3) with respect to t, one finds [14]:

$$\frac{dC}{dt}(t) = \frac{A}{\delta}\frac{e^{-\frac{(t-t_p)}{\delta}}}{\left(1 + \nu e^{-\frac{(t-t_p)}{\delta}}\right)^{\frac{\nu+1}{\nu}}} \tag{4}$$

The expression presented in (3) is an asymmetrical sigmoid, containing one acceleration and one deceleration phase. Therefore, it is adequate to describe one epidemiological wave.

When multiple epidemiological waves occur in the same region during the same disease outbreak, it is convenient to describe the accumulated number of infections as a sum of sigmoids [14]. In this case, the set of parameters A, t_p, ν, δ is generalized to $A_j, t_{p,j}, \nu_j, \delta_j$, where j is an integer value varying from 1 to the number N_s of sigmoids adopted to describe the disease behavior. The total number of infected people is then given by:

$$A = \sum_{j=1}^{N_s} A_j \tag{5}$$

The number of sigmoids N_s is estimated by analyzing the filtered second derivative of the observed data. A transition from a deceleration to an acceleration phase indicates the onset of a new epidemiological wave. The procedure is described in detail in our previous paper [14].

The estimation of the parameters $A_j, t_{p,j}, \nu_j, \delta_j$ is performed by using a standard numerical optimization technique to solve a constrained optimization problem, aiming to minimize the quadratic residue between the observed data and the

model output $C(t)$. To that end, the well-known Sequential Quadratic Programming algorithm [15] is adopted here. It is worth remarking that the parameters $A_j, t_{p,j}, \nu_j, \delta_j$ are estimated once again once more data is available. Thus, at each iteration they are constant, but their estimated values vary with time.

Let $\hat{A}(t)$ represent the estimation of the value of A performed with all available data until a given date t. It is important to point out that, in a model representing the outbreak of a disease in a given locality, the value of A is actually a constant value. The term that is a function of time is the estimation of such value. This estimation may vary according to unpredicted changes in the disease dynamics.

An analysis of the trend of $\hat{A}(t)$ along time provides an important information for health authorities. If this value is approximately constant for a long period, one may assume that the disease is under control and converging to a final value. On the other hand, if an increasing trend is observed, then further actions to contain the disease would be advised.

The calculation of $\hat{A}(t)$ requires repeating the model calibration for every day between specified starting and ending dates. That means that the determination of the value of $\hat{A}(t)$ in an interval range $t_a \leq t \leq t_b$ requires an estimation of the model parameters at each time instant $t = t_a, t_a + 1, ..., t_b$ with all available data until that specific date. This procedure may be very expensive computationally. That is the reason why we have chosen a model with a closed form, expressed in eq. (3). The use of a model that does not require solving a differential equation allows for a computationally efficient implementation.

III. RESULTS AND DISCUSSION

To perform this study, data describing the daily infections of the COVID-19 pandemic in Mexico until March 17th 2021 were downloaded from an official website of the Mexican government [16]. Then, four states and four cities were chosen randomly along the most populated ones.

Figure 1 presents the results of this study for the Mexican states of Distrito Federal, Guanajuato, Jalisco, and Veracruz. Each row in the figure corresponds to one of these states, which were sorted in alphabetical order. Three graphs are presented for each state: (a) the cumulative number of infections $C(t)$, comparing the observed data and the model output described in eq. (3); in the bottom right of each graph, the value of the root mean-squared error (RMSE) is presented; (b) the daily number of infections $dC(t)/dt$, comparing the raw observed data, the observed data filtered through a 7-day moving average filter, and the model output, as in eq. (4); and (c), the predicted final number of cases $\hat{A}(t)$.

Similar results are presented in Fig. 2 for the Mexican cities of Juarez, Leon, Puebla, and Tijuana.

From the results, it is clear that the technique is able to find a very good fit for the curve of the accumulated cases. In fact, the root mean-squared error RMSE value found in each case is very small, when compared to the number of cases. On the other hand, the data containing the daily number of infections presents a behavior similar
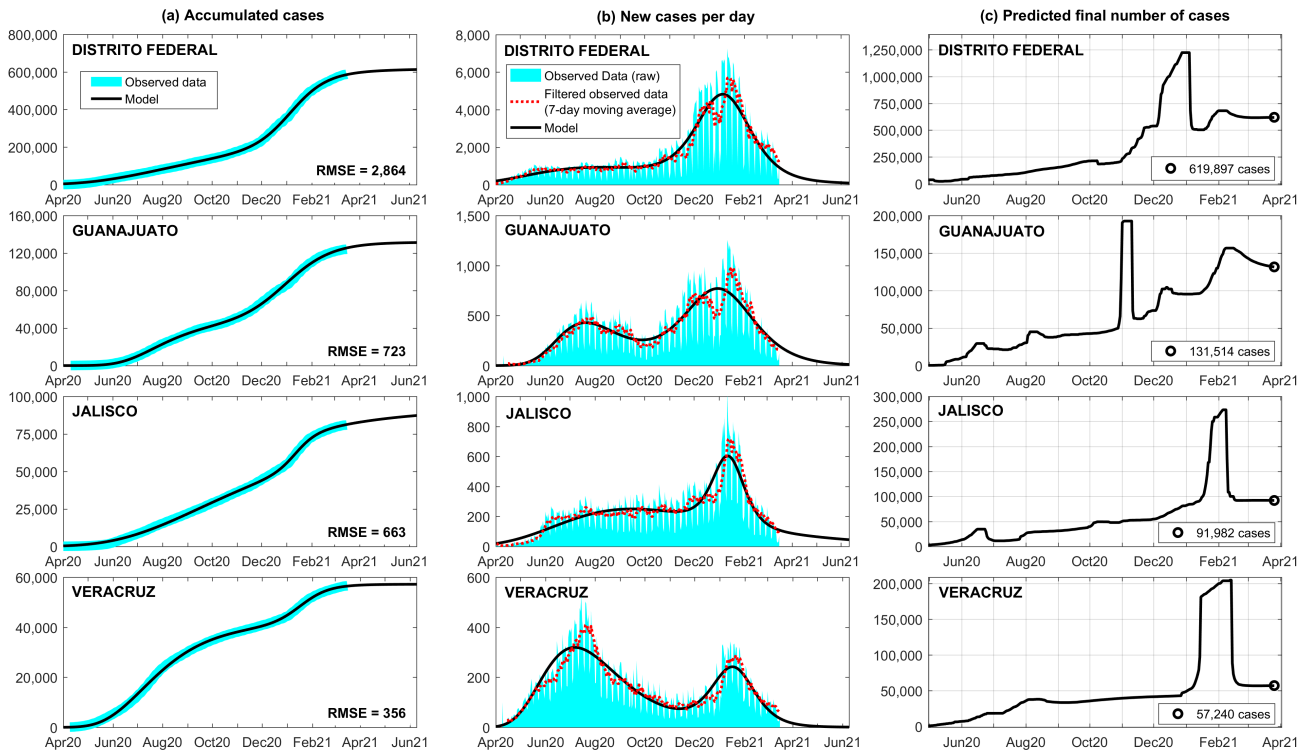
Fig. 1.   Results for the Mexican states of Distrito Federal, Guanajuato, Jalisco and Veracruz.

to a signal corrupted by high sensor noise, probably due to a non-uniform delay between the infection occurrence and its notification to health agencies. Despite this fact, the model can accurately approximate the general trend of the daily cases. This type of noisy behavior cannot be observed in the accumulated cases data because of its integrative nature, that balances out the fluctuations and gives it a robustness against noise. It should also be noted the similarity between the model output and the filtered daily data.

It can be seen that every state and city had two epidemiological waves and that the second wave had noticeable peaks at approximately the same period of time (Jan-Feb/2021). The only exception is Juarez, which had a pronounced peak around Nov/2020. The model output describing the daily infections shows that, in most cases, the second wave started around Oct/Nov 2020 and had more infections than the first wave. Only at the state of Veracruz the second wave did not surpass the number of cases of the first one. It is also evident that the model can successfully represent this multiple-wave behavior.

The graph with the predicted final number of cases expresses valuable information about the trend of the pandemic across the year in Mexico. It can be seen that, in the second semester of 2020, the graph is nearly constant for most of the states and cities. This behavior reveals a stabilization on the number of expected infections in those regions at that time, since the whole country was recovering from the peak in the middle of the year. Around December, however, this prediction rises suddenly, indicating the possible arrival of a second wave there. The later results confirmed the occurrence

of such second wave.

In the beginning of the second wave, the value of the predicted number of cases suffers a very large increase, but then decreases to a smaller value, in which the prediction has remained for the last four weeks of the analysis. Such value, however, is still significantly higher than the predictions performed in the second semester of 2020, which was expected, because the number of infected people is expected to be higher if there are two epidemiological waves instead of only one. Furthermore, the initially very large increase in the prediction during the onset of the second wave, followed by a later stabilization, indicates that the analysis is sensitive to the amount of data available and that it is capable to perform a better forecast when there is significant new data describing the new wave. On the other hand, the quality of the forecast may be inferred by analyzing its variations along time, as a forecast is intrinsically more reliable if it remains constant for several weeks. Such regular behavior, almost constant, also indicates that the epidemic is under control. This is the main conclusion that can be drawn from the analysis concerning the data of the more recent four weeks.

IV. CONCLUSIONS

This paper employed a classic epidemiological model to represent the dynamics of the COVID-19 in Mexican states and cities. A trend analysis approach was employed allowing to assess if the pandemic is already stabilized or requires a more strong action from the local governments. A model with a closed-form solution was adopted, in order to allow for a more computationally efficient implementation.
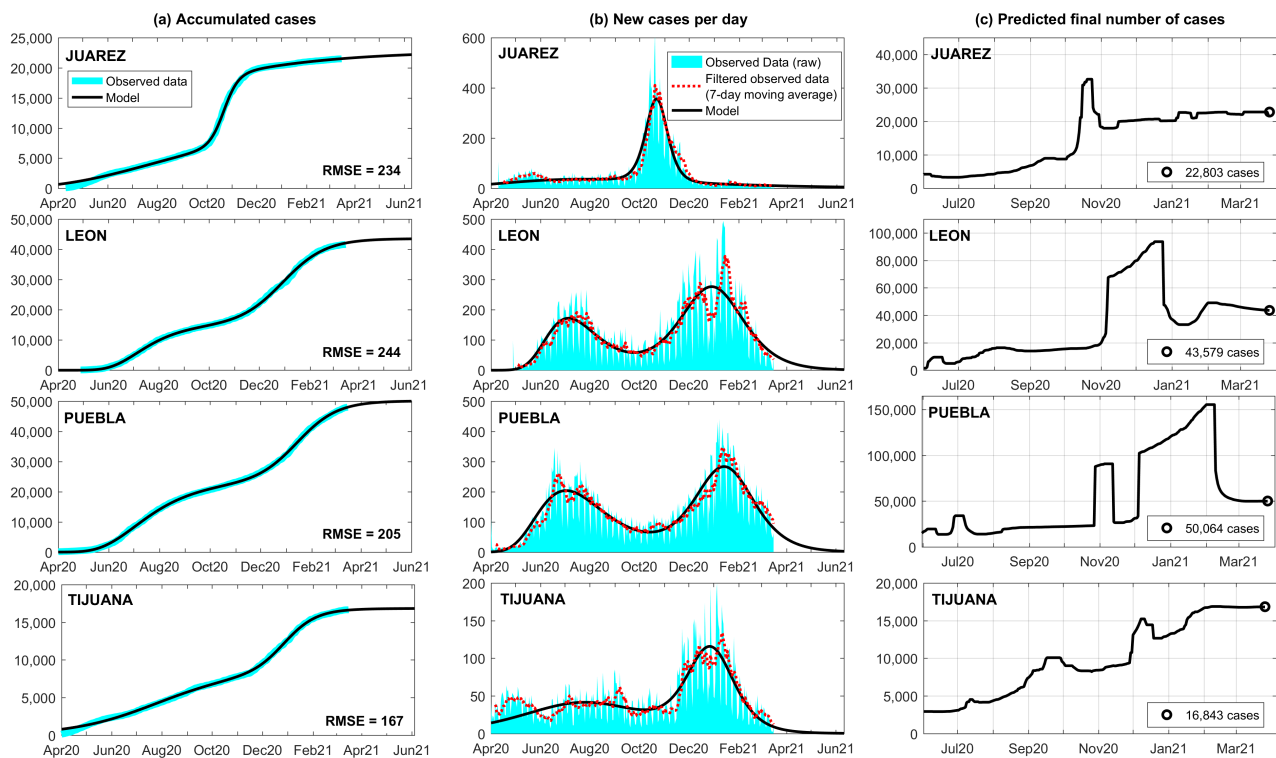
Fig. 2. Results for the Mexican cities of Juarez, Leon, Puebla and Tijuana.

The technique provides more useful information when applied for a city, instead of a state, because it allows the municipal authorities to take action according to the local behavior of the disease.

Future works could run the trend analysis with different model structures and extend it to other regions in the world. Furthermore, the same approach could be adopted to analyze the trend of other diseases.

REFERENCES

[1] C. A. Devaux, J.-C. Lagier, and D. Raoult, "New insights into the physiopathology of covid-19: Sars-cov-2-associated gastrointestinal illness," *Frontiers in medicine*, vol. 8, 2021.

[2] M. Correale, L. Tricarico, M. Fortunato, G. Dattilo, M. Iacoviello, and N. D. Brunetti, "Infection, atherothrombosis and thromboembolism beyond the covid-19 disease: what similar in physiopathology and researches," *Aging Clinical and Experimental Research*, pp. 1–6, 2021.

[3] G. Tsivgoulis, L. Palaiodimou, A. H. Katsanos, V. Caso, M. Köhrmann, C. Molina, C. Cordonnier, U. Fischer, P. Kelly, V. K. Sharma *et al.*, "Neurological manifestations and implications of covid-19 pandemic," *Therapeutic advances in neurological disorders*, vol. 13, p. 1756286420932036, 2020.

[4] D. Pradhan, P. Biswasroy, G. Ghosh, G. Rath *et al.*, "A review of current interventions for covid-19 prevention," *Archives of medical research*, 2020.

[5] X. Jiang, J. Hwang, D. V. Shah, S. Ghosh, and M. Brauer, "News attention and social-distancing behavior amid covid-19: How media trust and social norms moderate a mediated relationship," *Health Communication*, pp. 1–10, 2021.

[6] S. A. Abdullah, S. Owyed, A.-H. Abdel-Aty, E. E. Mahmoud, K. Shah, and H. Alrabaiah, "Mathematical analysis of covid-19 via new mathematical model," *Chaos, Solitons, and Fractals*, vol. 143, p. 110585, 2021.

[7] A. Zeb, E. Alzahrani, V. S. Erturk, and G. Zaman, "Mathematical model for coronavirus disease 2019 (covid-19) containing isolation class," *BioMed research international*, vol. 2020, 2020.

[8] H. M. Paiva, R. J. M. Afonso, I. L. de Oliveira, and G. F. Garcia, "A data-driven model to describe and forecast the dynamics of covid-19 transmission," *PloS One*, vol. 15, no. 7, p. e0236386, 2020.

[9] I. Ibarra-Nava, J. A. Cárdenas-De La Garza, R. E. Ruiz-Lozano, and R. G. Salazar-Montalvo, "Mexico and the covid-19 response," *Disaster Medicine and Public Health Preparedness*, vol. 14, no. 4, pp. e17–e18, 2020.

[10] V. Suárez, M. S. Quezada, S. O. Ruiz, and E. R. De Jesús, "Epidemiología de covid-19 en méxico: del 27 de febrero al 30 de abril de 2020," *Revista clinica espanola*, vol. 220, no. 8, pp. 463–471, 2020.

[11] M. R. Tocto-Erazo, J. A. Espíndola-Zepeda, J. A. Montoya-Laos, M. A. Acuña-Zegarra, D. Olmos-Liceaga, P. A. Reyes-Castro, and G. Figueroa-Preciado, "Lockdown, relaxation, and acme period in covid-19: A study of disease dynamics in hermosillo, sonora, mexico," *PloS one*, vol. 15, no. 12, p. e0242957, 2020.

[12] A. Smirnova and G. Chowell, "A primer on stable parameter estimation and forecasting in epidemiology by a problem-oriented regularized least squares algorithm," *Infectious Disease Modelling*, vol. 2, no. 2, pp. 268–275, 2017.

[13] F. Richards, "A flexible growth function for empirical use," *Journal of Experimental Botany*, vol. 10, no. 2, pp. 290–301, 1959.

[14] H. M. Paiva, R. J. M. Afonso, F. M. S. L. Alvarenga, and E. A. Velasquez, "A computational tool for trend analysis and forecast of the Covid-19 pandemic," *Applied Soft Computing*, p. 107289, 2021.

[15] P. E. Gill and E. Wong, "Sequential quadratic programming methods," in *Mixed integer nonlinear programming*. Springer, 2012, pp. 147–224.

[16] Mexican Government, "Covid-19 México," https://datos.covid-19.conacyt.mx/index.php#DownZCSV, accessed: March 18th, 2021.