Modeling and prediction of COVID-19: a q-exponential model approach

Freddy H. Marín-Sánchez, Jhon Alexanders Miranda Echeverri, Pablo Alberto Osorio Marulanda, Santiago Isaza Cadavid

Abstract

In this work we proposed a deformed q-exponential model to represent and forecast the dynamic behavior of the curve of Covid-19 reported cases for 7 different countries. This model is of the textural or univariate type with three dynamic parameters estimated at every point of the time. The model has an easy computational implementation and results for the forecast are graphically illustrated under two scenarios, the first one for a daily and ten days forecast for Colombia, Mexico, India, and Argentina and the second one for five days forecast of Italy, Spain, and Germany using moving windows.

Keywords: Covid-19 modelling, Covid-19 prediction, q-exponential function, moving windows

1. Introduction

The coronavirus, caused by the virus known as SARS-CoV-2 detected in Wuhan, China, on December 31, 2019 has changed the world as we know it; its effects on the respiratory system make its propagation very easy and fast, resulting in a large number of deaths and infected worldwide. The high rate of spread of this virus and the inconvenience it has brought to public health prompted the World Health Organization (WHO) to announce on January 31, 2020 the COVID-19 epidemic as a public health emergency of international concern; the same health agency classified the emergency as a pandemic on March 11, 2020 [1].

In the midst of this situation, many scientific studies have been carried out to propose solutions to all the effects that the pandemic has generated in the fields of health, unemployment, security, food shortage, and various social problems derived from it ([2], [3], [4]). This solutions include both simple models and highly complex mathematical systems. From the epidemiological point of view, it is of vital importance to model and predict the behaviour of the variables that help explain the dynamics of this pandemic; for example some of them are the reproductive number cash (ρ_0), which represents the number of people infected by a person, morbidity, which states how many new cases there are every day and how many cumulative, lethality, which determine the percentage of people who have died with respect to cases identified as positive and the number of patients in Intensive Care Units (ICU), among others. While the measurement of these variables may be under-reporting (it can be known how many patients have tested positive depending on the number of tests made), all of them can be estimated with inaccurate results due to high uncertainty that may be present in such measurements. For this reason, one alternative to measure them, with the limitations that this represents, is to use the daily data published by the different governments or health entities of each country or region.
Since the beginning of the pandemic, many models have been proposed that attempt to explain or forecast the most relevant variables using different approaches. Some proposals have taken up the classic approaches of dynamic systems with some modifications. You can also find works developed with artificial intelligence, neural networks [5] and even stochastic differential equations[6][7] ; others have used statistical, econometric and time series methods, some combinations with various methodologies can even be found in recent literature.

Email addresses: `fmarinsa@eafit.edu.co` (Freddy H. Marín-Sánchez), `jmirand1@eafit.edu.co` (Jhon Alexanders Miranda Echeverri), `paosoriom@eafit.edu.co` (Pablo Alberto Osorio Marulanda), `sisazac@eafit.edu.co` (Santiago Isaza Cadavid)

Under the approach of differential equations and dynamic systems, mathematical models of the SIR type (Susceptible-Infected-Recovered) can be highlighted, with various added values, which seek to represent the dynamics of the virus, making possible to predict new cases in a region and determined period. The SIR model has also been studied with stochastic parameters [5] and other derivatives that include the classic SEIR (Susceptible-Exposed-Infected-Recovered) model, without quarantine controls [5] and even a model with a simulation of drastic containment measures [8]. The SIRU model with a system of ordinary differential equations [9] and others that consider growth behavior after mitigation policies, removing the susceptible ones from the transmission process and the elimination of symptomatic infected is counted. Finally the system dynamics is governed by a system of differential equations [10, 1].
Other works are related to exponential or second derivative models such as the one proposed in [11], which uses a technique to characterize the coronavirus epidemic in China with accumulated cases during the first two months. Subsequently, they improved the analysis with an exponential model under the assumption of close-population and concluded that the epidemic has a non-linear and chaotic behavior, which responds highly to the actions that are implemented in the population, such as social distancing. Meanwhile in [12] they model the fatality curves of the COVID-19 disease, using the accumulated number of deaths and study the effectiveness of possible intervention strategies applying the Richards growth model (RGM) and a q-exponential function to the lethality curves for various countries.

On the other hand, in the models that have a perspective more focused on statistical methods is the proposal of [13] in which they apply a GLM (Generalized linear model), the Richards model and a sup-epidemic model, estimating the best solution fit by minimizing the sum of squared errors of the data and then using parametric Bootstrap to quantify the uncertainty of the results of each solution. More approximations that at first glance might seem less ideal because they are simpler to represent the spread of the virus, have also obtained favorable results when compared with data provided by governments or certified organizations. For example, [14] shows a quantitative description of the epidemic in China and validates it with data collected from Austria up to that point and also [15] using traditional time series prediction models.

Undoubtedly, the modeling of an epidemiological system such as that of COVID-19 can have a high degree of complexity and high computational cost. Additionally, it can present some problems when the variables must be applied in real time, moreover if there is little data and the parameters carry with them a large part of the uncertainty of the system. For this reason it can be very useful to model the variables separately, providing to each variable of a univariate mathematical representation independent of the others.It can be seen in models such as the one presented in ([16]) to forecast the number of infections in Mexico and the work of ([15]) to model the accumulated number of infected persons recovered and deceased globally, adopting a time series approach combined with exponential smoothing. These variables are relevant to prepare the medical equipment and the physical infrastructure of clinics, hospitals and cemeteries so that more deaths can be prevented, the disease can be correctly fought. Other associated problems could be controlled, as with the evolution of the epidemic in the world having daily forecasts and more days ahead can help implement public policies that seek to reduce the effects of the pandemic in areas with high risk of spreading the virus.

Given the importance of knowing the possible number of infections in the future, in this work we want to propose a univariate temporal model of the q-exponential type with 3 dynamic parameters to explain the behavior of daily reported cases of Covid-19. This model is easy to implement computationally, highly efficient, it is updated with the information of each day, it allows to make daily forecasts of growing curves with a range of up to 10 days and forecasts of 5 days ahead with moving windows for curves in their phase of flattening, for any country or region with daily information on the number of cases; it also shows very low relative forecast errors.
In the development of this work, forecasts were made of up to 10 days of the number of total confirmed cases from the beginning of the pandemic until July 31 in Colombia, Mexico, Argentina and India, using the q-exponential model with dynamic parameters and for the forecast of up to 5 days in Italy, Spain and Germany, mobile windows based on q-exponential and fixed parameters were used. The results presented in this work are calculated according to the data reported for the daily number of confirmed cases by the Colombian Ministry of Health [17] and the Johns Hopkins University of Medicine [18] for the other countries.

This paper is organized as follows: In the second section the q-exponential functions are described, the q-exponential model is built and the parameter estimation technique is presented. A brief description of the data and forecast results for different countries is shown in the third section. The last section is dedicated to the conclusions.

2. q-exponential model and parameters estimation

Definition 1. An exponential function with respect to the parameter q or q-exponential function, for a variable x is defined as:

$$exp_q(x) = e_q^x :\equiv \begin{cases} [1+(1-q)x]^{\frac{1}{1-q}} & \text{if } 1+(1-q)x > 0 \\ 0 & \text{if } 1+(1-q)x \leq 0 \end{cases}$$

Where $x, q \in \mathbb{R}$. Note that $e_1^x :\equiv \lim_{q\to1^+} e_q^x = \lim_{q\to1^-} e_q^x = e^x \quad \forall x$.

Furthermore, if $q < 1$ the q-exponential function fades for $x \leq \frac{1}{q-1}$ and when x increases from $\frac{-1}{1-q}$ to ∞, it continuously and monotonously increases from 0 to ∞.
On the other hand, if $q > 1$ the q-exponential function turns out to be divergent for $x > \frac{1}{q-1}$ and when it increases from $-\infty$ to $\frac{1}{q-1}$, it increases continuously and monotonously from 0 to ∞.
The figures 1(a) and 1(b) show the behavior of a q-exponential function for different values of q with $x \in [0, 20]$. The properties of the q-exponential function and other details can be consulted in [19, 20, 21].

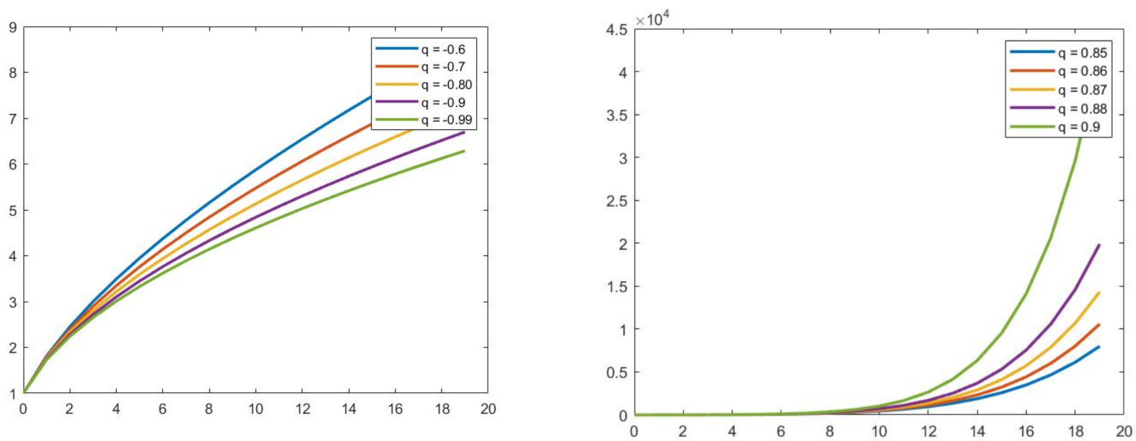


Figure 1: (a) q-exponential function with negative values of q. (b) q-exponential function with negative values of q

2.1. q-exponential model

2.1.1. Basic Assumptions

The dynamic behavior of the number of daily infections reported of Covid-19 for an infected geographic region under the following postulates:

- Information about daily reports number is reliable.
- All recovered patients can be infected again.
- Virus detection tests are performed daily and are reported the day the results are obtained.

The proposed model is of the textural or univariate type, which means that only the daily historical data of reported infections is considered and excludes other dynamics of the classic epidemiological models associated

with deaths, exposed, isolated, etc, such as the SEIR model with stochastics parameters, SEIR without quarantine and SIRU ([6],[5],[9]). So the statistical inference and the estimation of the parameters are extracted exclusively from the data of the available sample.
Since the q-exponential function is continuous and the observed data for contagions is in discrete time, the approximation of the reported contagion curve is performed at discrete points of the q-exponential function.

2.1.2. Model building

Suppose there are constants $f(\tau_n)$, $q(\tau_n)$ y $m(\tau_n)$ that explain "better" the behavior of the curve $C(\tau_t)$ up to time τ_n using the expression

$$C(\tau_t) \approx f(\tau_n) exp_{q(\tau_n)}(m(\tau_n)\tau_t); \quad C(\tau_0) = C(0) \quad \text{is a constant} \tag{1}$$

For observation $C(\tau_{n+1})$ we would have new parameters $f(\tau_{n+1})$, $q(\tau_{n+1})$ and $m(\tau_{n+1})$ that would explain the contagion curve updated up to time τ_{n+1} and so on, so that for each new observation the curve $C(\tau_t)$ is updated.The figure 2 illustrates the approximation of a curve of infections reported using a q-exponential for dynamic behavior for infection curve for 10, 15, 20 and 25 days.

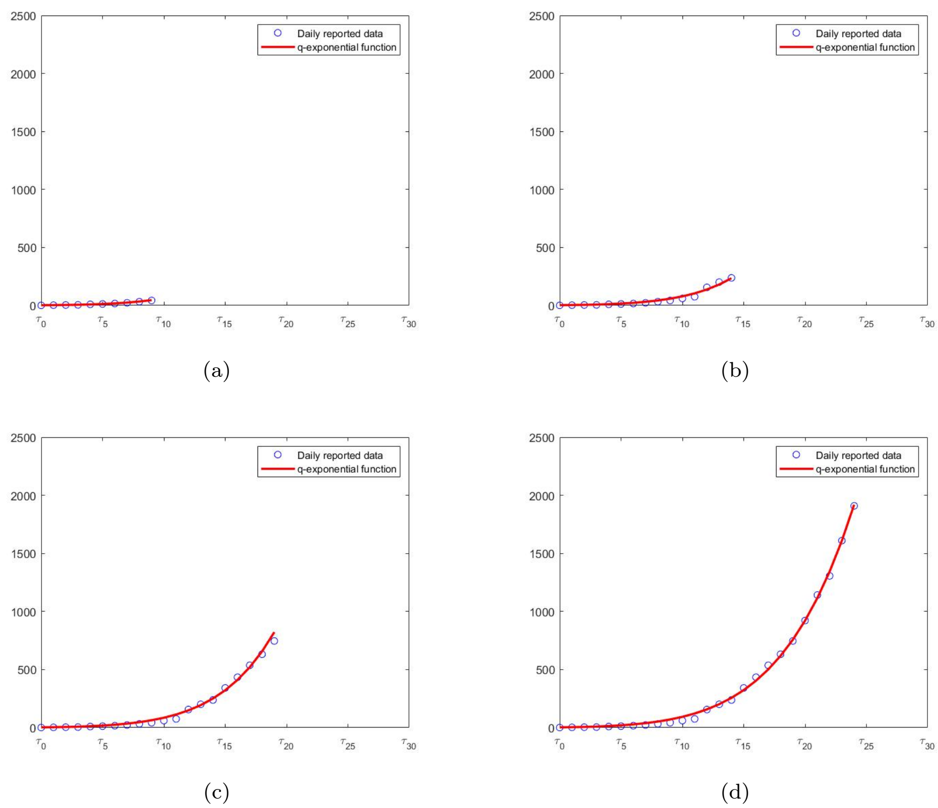


Figure 2: Approximation of reported cases curve using q-exponential functions for 10, 15, 20 and 25 days

2.2. ***Parameters estimation***

Consider the q-exponential model (1) with the initial condition $C(\tau_0) = C(0)$ and an unknown parameters vector $\Theta = [f(\tau_t), m(\tau_t), q(\tau_t)]$; $t = 1, 2, \cdots, n$. Consider the finite partition of the time interval $[0, T]$, $0 = \tau_0 \leq \tau_1 \leq \ldots \leq \tau_{n-1} \leq \tau_n = T$ with $\delta t = \tau_t - \tau_{t-1} \geq 0$; where $t = 0, 1, 2, \ldots, n$ and a daily contagion curve $C(\tau_t) \in [0, T]$ observed at discrete points for the times $\{\tau_0, \tau_1, \ldots, \tau_{n-1}, \tau_n\}$. The objective is to estimate the unknown parameters vector Θ given the historical data of the sample $C = [C\tau_0, C\tau_1, \cdots C\tau_n]$ with equally spaced data.

The estimation of the number of total cases $C(\tau_n)$ for a day τ_n is given by its estimation $\hat{C}(\tau_n)$ in the q-exponential model

$$\hat{C}(\tau_n) = \hat{f}(\tau_n)(1 + (1 - \hat{q}(\tau_n))\hat{m}(\tau_n)\tau_n)^{\frac{1}{1-\hat{q}(\tau_n)}} \tag{2}$$

If we extend the last equation to the interval $[\tau_0, \tau_n]$, the difference between real reported data and estimated data is given by the Sum of Squared Errors $(SSE) = g(\epsilon_t)$; where $\epsilon_t = C(\tau_t) - \hat{C}\tau_t$.

$$\begin{aligned} g(\epsilon_t) &= \sum_{i=0}^{n} \Big(C(\tau_t) - \hat{C}(\tau_t)\Big)^2 && (3) \\ &= \sum_{i=0}^{n} \Big(C(\tau_t) - \hat{f}(\tau_n)(1 + (1 - \hat{q}(\tau_n)\hat{m}(\tau_n)\tau_t)^{\frac{1}{1-\hat{q}(\tau_n)}}\Big)^2 && (4) \end{aligned}$$

If $g(\epsilon_t)$ has a minimum it happens for values of $\hat{f}$, $\hat{m}$ and $\hat{q}$ that satisfy the equation

$$\frac{\partial g(\epsilon_t)}{\partial \Theta} = \vec{0} \qquad \text{i.e.,} \quad \hat{\Theta} = \underset{\Theta}{Arg}\ min(g(\epsilon_t)) \tag{5}$$

To calculate these derivatives let us consider the following auxiliary variables

$$v_t = 1 + (1 - \hat{q}(\tau_n))\hat{m}(\tau_n)\tau_t \tag{6}$$

$$w_t = \frac{\ln(v_t)}{1 - \hat{q}(\tau_n)} \quad (^1) \tag{7}$$

The parameter estimation $\hat{f}$ $\hat{m}$ and $\hat{q}$ are given by the solution of (5) considering the next partial derivatives:

$$\begin{aligned} \frac{\partial g(\epsilon_t)}{\partial \hat{f}} &= \frac{\partial}{\partial \hat{f}} \sum_{t=0}^{n} \epsilon_t^2 \\ &= -\sum_{t=0}^{n} 2\epsilon_t (v_t)^{\frac{1}{1-\hat{q}(\tau_n)}} \\ &= -\sum_{t=0}^{n} 2(C(\tau_t) - \hat{f}(\tau_n)(v_t)^{\frac{1}{1-\hat{q}(\tau_n)}})(v_t)^{\frac{1}{1-\hat{q}(\tau_n)}} && (8) \end{aligned}$$

$$\begin{aligned} \frac{\partial g(\epsilon_t)}{\partial \hat{m}} &= \frac{\partial}{\partial \hat{m}} \sum_{t=0}^{n} \epsilon_t^2 \\ &= \sum_{t=0}^{n} -2\epsilon_t \hat{f}(\tau_n) \left(v_t^{\frac{q(\hat{\tau_n})}{1-q(\hat{\tau_n})}} \tau_t \right) \\ &= \sum_{t=0}^{n} (-2\hat{f}(\tau_n)(v_t)^{\frac{\hat{q}(\tau_n)}{1-\hat{q}(\tau_n)}})(C(\tau_t) - \hat{f}(\tau_n)(v_t)^{\frac{1}{1-\hat{q}(\tau_n)}})\tau_t && (9) \end{aligned}$$

[1] ln which appears in w_t of equation (7) is not calculated in the sense of [20]; it is calculated in the classical sense.

$$
\begin{aligned}
\frac{\partial g(\epsilon_t)}{\partial \hat{q}} &= \frac{\partial}{\partial \hat{q}} \sum_{t=0}^{n} \epsilon_t^2 \\
&= \sum_{t=0}^{n} -2\epsilon_t \hat{f}(\tau_n) v_t^{\frac{1}{1-\hat{q}(\tau_n)}} \frac{1}{(1-\hat{q}(\tau_n)^2} \left(\frac{\partial}{\partial \hat{q}(\tau_n)} (1-\hat{q}(\tau_n)) \ln(v_t) - \frac{\partial}{\partial \hat{q}(\tau_n)} \hat{q}(\tau_n) \ln(v_t) \right) \\
&= \sum_{t=0}^{n} -\frac{2}{(1-\hat{q}(\tau_n))^2} u_i \hat{f}(\tau_n) v_t^{\frac{1}{1-\hat{q}(\tau_n)}} \left(\frac{(1-\hat{q})m(\tau_n)}{v_t} + \ln(v_t) \right) \qquad (10)
\end{aligned}
$$

2.3. Iterative procedure

An iterative procedure for the parameters vector $\hat{\Theta}(\tau_n) = (\hat{f}, \hat{m}, \hat{q})$, with the initial condition $\hat{\theta_0} = (\hat{f}_0, \hat{m}_0, \hat{q}_0)$ is realized. Define $\mathbf{J}_t = \frac{\partial \hat{C}_{\tau_t}}{\partial \Theta}$ and the first order estimate given by a new estimate $\theta + \delta$

$$
SSE(\theta + \delta) \approx \sum_{t=0}^{n} (C_{\tau_t} - \hat{C}_{\tau_t} - \mathbf{J}_t \delta)^2 \qquad (11)
$$

Taking the derivative respect to δ and setting to zero, in vectorial form

$$
\mathbf{J}^T \mathbf{J} \delta = J^T [C_{\tau_t} - \theta)] \qquad (12)
$$

Iterative step given a damping factor λ

$$
(\mathbf{J}^T J + \lambda I)\delta = J^T [C_{\tau_t} - \hat{C}_{\tau_t}(\theta)] \qquad (13)
$$

Daily forecasts are made by estimating the parameters $f(\tau_n)$, $q(\tau_n)$ and $m(\tau_n)$ of the daily contagion curve with the data reported up to time τ_n using the estimation technique described in the section (3.2); to obtain greater precision in the forecast, it is very useful to calibrate the parameter $\hat{f}(\tau_n)$ by adjusting it to the last data in the sample using the expression:

$$
\hat{\hat{f}}(\tau_n) = \frac{C(\tau_n)}{exp_{\hat{q}(\tau_n)}(\hat{m}(\tau_n)\tau_n)} \qquad (14)
$$

So that the forecast for the next day

$$
\hat{C}(\tau_{n+1}) = \hat{\hat{f}}(\tau_0) exp_{\hat{q}(\tau_n)}(\hat{m}(\tau_n)\tau_{n+1}) \qquad (15)
$$

has a lower relative error.
When the curve of reported infections is in its growth phase (before some flattening), forecasts can be made for several days ahead using the expression $\hat{C}(\tau_{n+k}) = \hat{\hat{f}}(\tau_n) exp_{\hat{q}(\tau_n)}(\hat{m}(\tau_n)\tau_{n+k})$ for $k = 1, 2, \cdots, p$.

Estimations and predictions are subject to errors in the fitted model, parameter deviations and external uncertainties. That is why an interval of values is required instead of a punctual one so that given a confidence level between 0 and 1, the possible values that might take a curve of cases at a specific time are contained with that probability. The procedure to generate confidence bands for predictions is described below.

- For a time τ_{n+k}, find the prediction $\hat{C}_{n+k} = \hat{C}(\tau_{n+k})$.
- Specify a confidence level.
- Compute the SSE using equation (3), the degrees of freedom as DOF$= \tau_{n+k} - p$ where p is the number of estimated parameters (3 for this case) and the inverse of the covariance matrix of the parameters R.

- Set RMSE= $\sqrt{\frac{\text{SSE}}{\text{DOF}}}$
- Get the derivative with respect to parameters at the value τ_{n+k} and find the numerical gradient $\mathbf{d}$ and set $\text{E} = \mathbf{d}R$.
- Compute the critical value using a T distribution as crit$= -t_{(1-\text{level})/2,\text{DOF}}$
- Compute the amplitude of the band $b = (\sqrt{\sum E \odot E})(\text{RMSE}(\text{crit}))$
- For the estimation $\hat{C}(\tau_n)$ the confidence band is $[\hat{C}_{\tau_{n+k}} - b, \hat{C}_{\tau_{n+k}} + b]$

2.4. Description of moving windows

It is possible to estimate advanced stages of the curve using a sample that gives priority to new available data. In other words, for a day τ_n with T observations, a time window w is specified to estimate with data of the last $T - w$ days. The value of w is between 4 (to ensure valid degrees of freedom) and T. Once the value for w is defined, the estimation and prediction follow the steps explained in Sections 3.2 and 3.3. The number of predictions can be extended up to 5 days. A good choice of the moving window should take into consideration the phases of the growth curve. In the first phase, the value of w is near to T because the curve has an accelerated behavior. During the second phase, the size might be around 7 and 15 days given a reduced growth rate. While for the third phase of curve flattening the size might vary between 4 and 21 days. Most of the data are updated daily and hence it can not be explicitly established at which stage is a curve of cumulative cases. To deal with this issue, exploration algorithms as Grid Search can be used to find the appropriate size of the time window. This requires an objective function based on estimations or predictions: Mean Squared Error (MSE), the amplitude of confidence bands or proximity to a statistic (mean, median, standard deviation).

3. Model Validation and prediction results

3.1. Data description

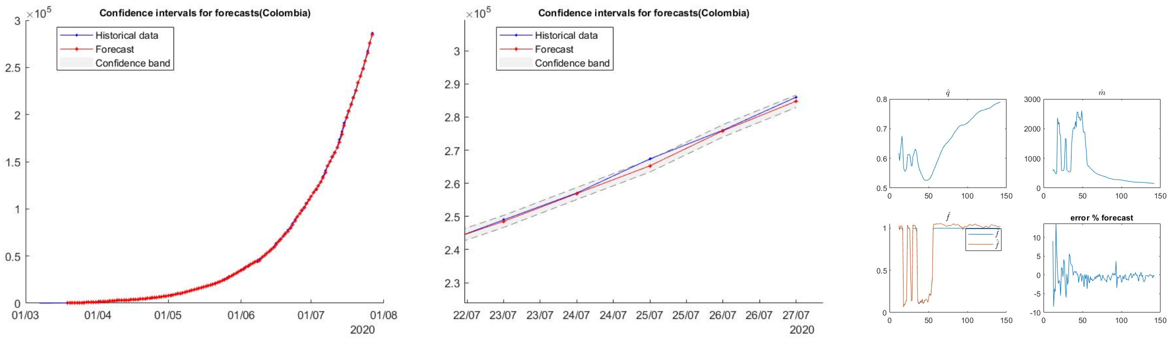
The data used to evaluate all the models proposed in this paper comes from the Minister of Health for Colombia and from Johns Hopkins University data for other countries such as Argentina, Germany, India, Italy, Mexico and Spain. The data presented figures out the number of total COVID-19 reported cases. The number of reported cases corresponds to those cases that were successfully detected, confirmed and reported, which means that the real number of cases must be higher than the number of cases which is reported.

3.2. Modeling and prediction based on recollecting all the historical data

The next results are based in a daily report and forecast for ten days to Colombia, Mexico, India and Argentina from the day of the first reported case (in every country) to 27, 23, 30 and 28 of July of 2020, respectively. The daily report takes the historical data from all the days behind for each iteration towards all the data. To induce the q-exponential model the first 12 reported data were taken to the first daily forecast. The forecast takes all the data except for the last ten days (in order to do error calculation and verify the model). In order verify the procedure R-squared, R-squared adjusted, the mean of relative error and the standard deviation is found for every country in every day forecast model in the tables 1, 2, 3, 4.

3.2.1. Colombia

The total number of reported cases of Colombia were taken from INS (Instituto Nacional de Salud) [17] from March 6 to July 27 of 2020 and the daily forecast are from March 18 to July 27.



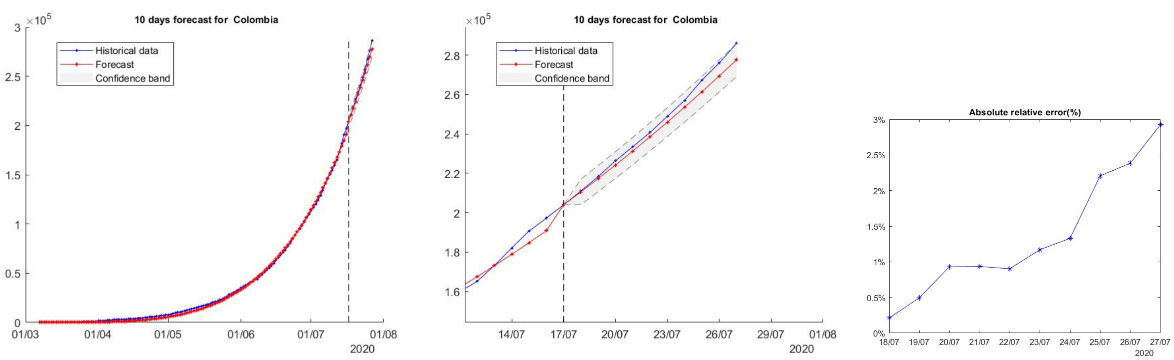
(a) (b) (c)

Figure 3: (a) Comparison between historical data and daily forecast of reported cases using model (2). (b) Detail of the daily forecast and confidence band. (c) Dynamic behavior of the estimator $\hat{f}, \hat{\hat{f}}, \hat{m}, \hat{q}$ and the absolute relative error for every day.

Country	R^2	R^2 adjusted	Mean relative error	Standard deviation(%)
Colombia	**0.99994**	**0.99997**	**-0.003657**	**0.02237**

Table 1

- Forecast for 10 days For ten days forecast the data was taken from march 6 to July 17 of 2020.



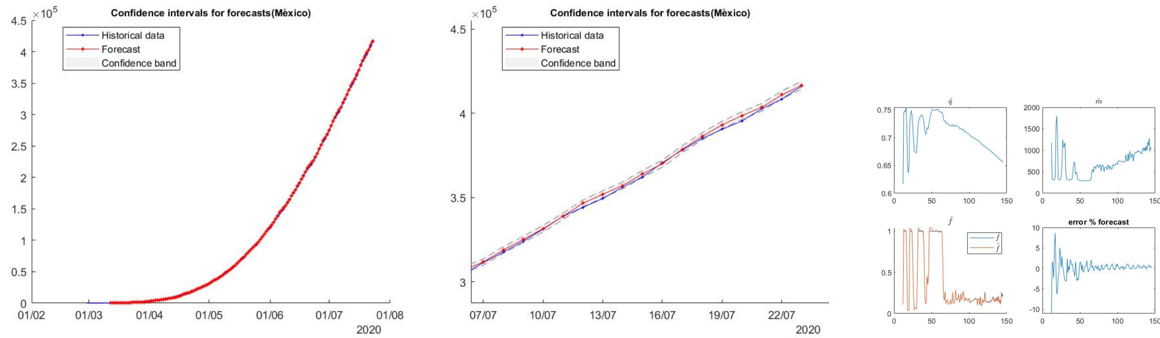
(a) (b) (c)

Figure 4: (a) Comparison/Validation between forecast for 10 days and reported historical data using model (2). (b) Detail of the 10 days forecast and confidence band. (c) Absolute relative error of the forecasts.

Figure 5

3.2.2. Mexico

The total number of reported cases of Mexico were taken from Johns Hopkins University from February 29 to July 23 of 2020 and the daily forecast are from March 12 to July 23.



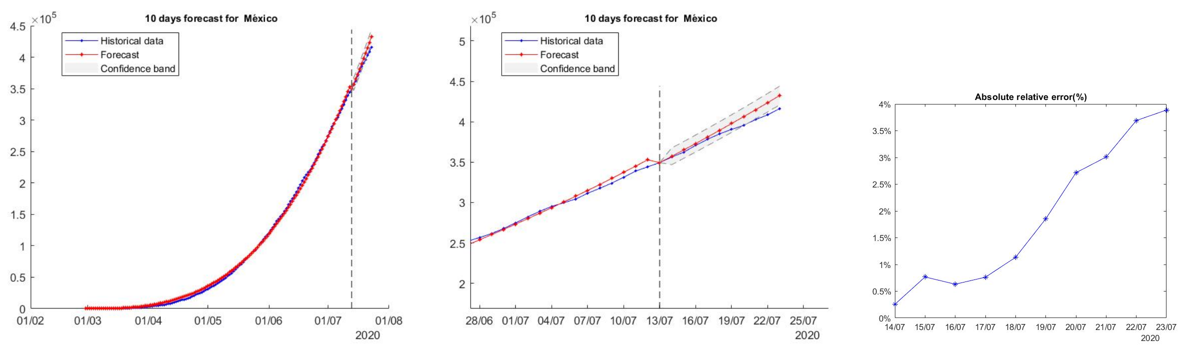
(a) (b) (c)

Figure 6: (a) Comparison between historical data and daily forecast of reported cases using model (2). (b) Detail of the daily forecast and confidence band. (c) Dynamic behavior of the estimator $\hat{f}, \hat{\hat{f}}, \hat{m}, \hat{q}$ and the absolute relative error for every day.

Country	R^2	R^2 adjusted	Mean relative error	Standard deviation(%)
Mexico	**0.99998**	**0.99998**	**0.001818**	**0.01632**

Table 2

- Forecast for 10 days For ten days forecast the data was taken from February 29 to July 13 of 2020.

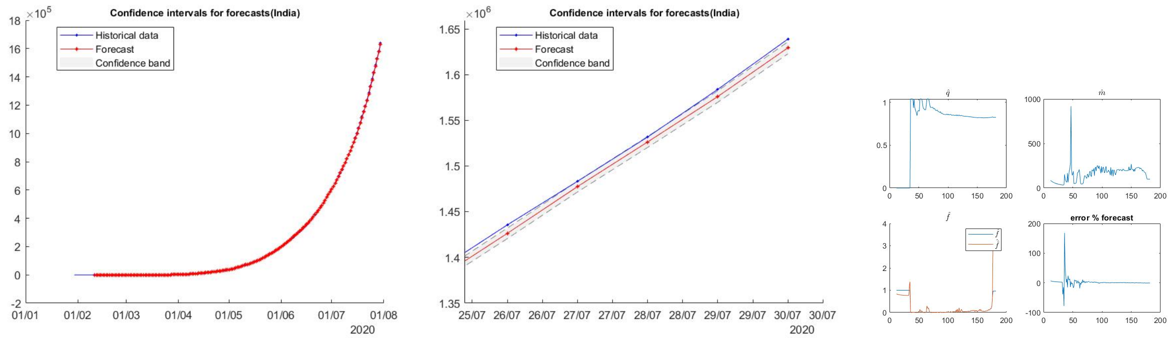


(a) (b) (c)

Figure 7: (a) Comparison/Validation between forecast for 10 days and reported historical data using model (2). (b) Detail of the 10 days forecast and confidence band. (c) Absolute relative error of the forecasts.

3.2.3. India

The total number of reported cases of India were taken from Johns Hopkins University from January 30 to July 30 of 2020 and the daily forecast are from February 11 to July 30.



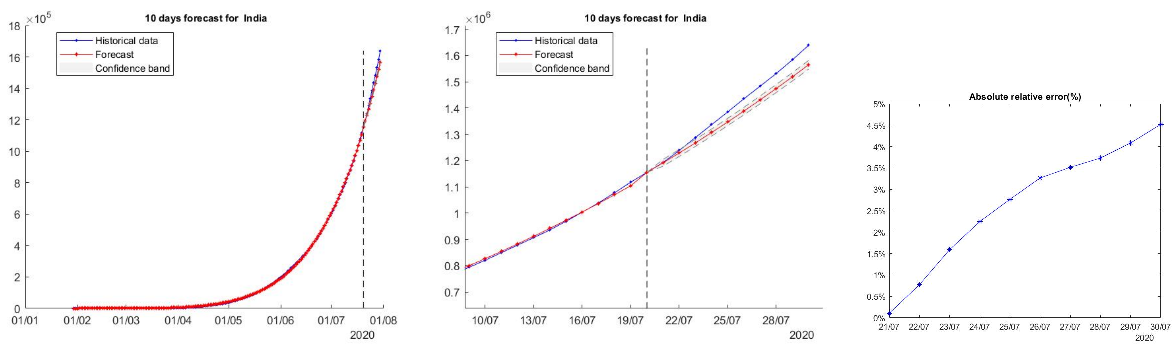
(a) (b) (c)

Figure 8: (a) Comparison between historical data and daily forecast of reported cases using model (2). (b) Detail of the daily forecast and confidence band . (c) Dynamic behavior of the estimator $\hat{f}, \hat{\hat{f}}, \hat{m}$,$\hat{q}$ and the absolute relative error for every day.

Country	R^2	R^2 adjusted	Mean relative error	Standard deviation(%)
India	**0.99998**	**0.99999**	**0.01793**	**0.1586**

Table 3

- Forecast for 10 days For ten days forecast the data was taken from January 30 to July 30 of 2020.

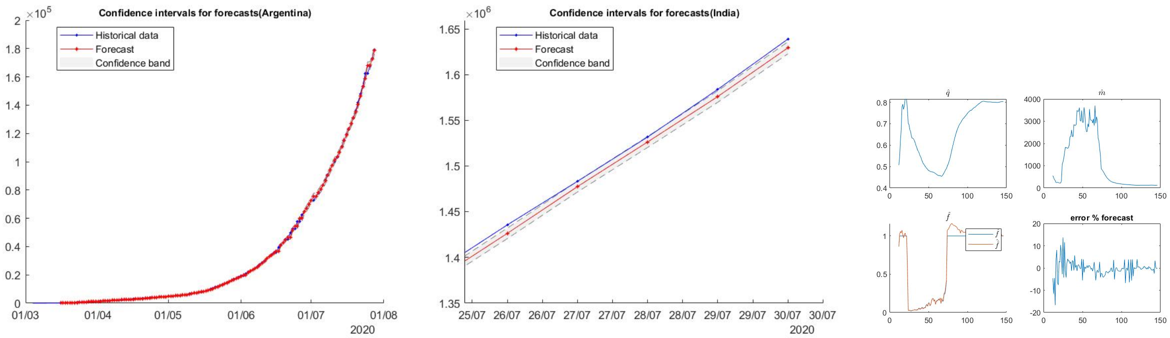


(a) (b) (c)

Figure 9: (a) Comparison/Validation between forecast for 10 days and reported historical data using model (2). (b) Detail of the 10 days forecast and confidence band. (c) Absolute relative error of the forecasts.

3.2.4. Argentina

The total number of reported cases of Argentina were taken from Johns Hopkins University from March 4 to July 28 of 2020 and the daily forecast are from March 16 to July 28.



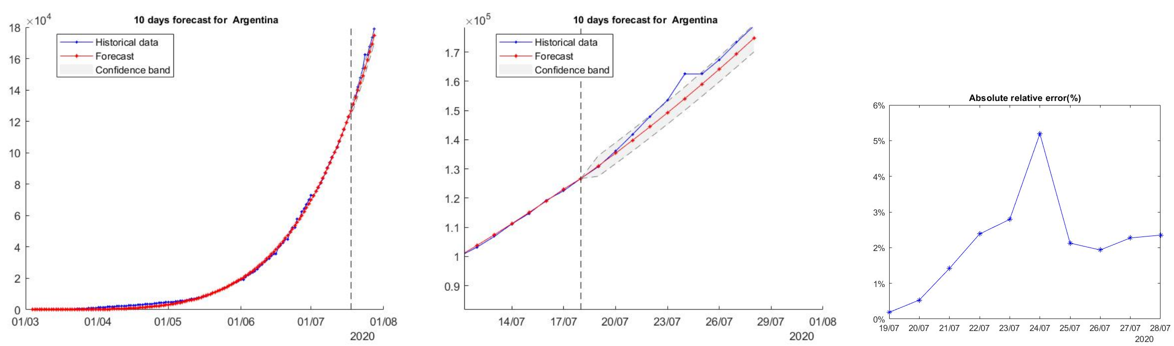
(a) (b) (c)

Figure 10: (a) Comparison between historical data and daily forecast of reported cases using model (2).(b) Detail of the daily forecast and confidence band . (c) Dynamic behavior of the estimator $\hat{f}, \hat{\hat{f}}, \hat{m}$,$\hat{q}$ and the absolute relative error for every day.

Country	R^2	R^2 adjusted	Mean relative error	Standard deviation
Argentina	**0.99962**	**0.99981**	**-0.002371**	**0.03653**

Table 4

- Forecast for 10 days For ten days forecast the data was taken from March 4 to July 18 of 2020. The day 24 of July the data base used for the research reported 0 new cases of infection for this day; it is supposed to be a mistake even thought affects the result of the forecast.



(a) (b) (c)

Figure 11: (a) Comparison/Validation between forecast for 10 days and reported historical data using model (2). (b) Detail of the 10 days forecast and confidence band. (c) Absolute relative error of the forecasts.

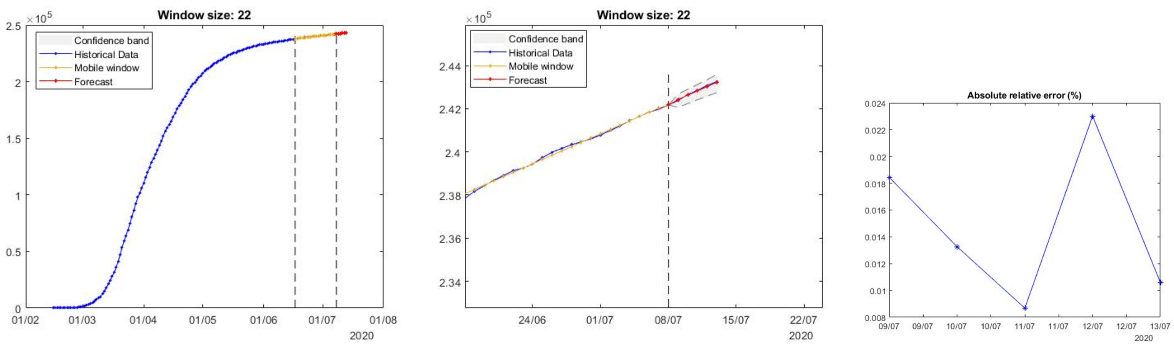
3.3. Modeling and prediction based on moving windows

The next results are based on a daily report and forecast for five days to Italy, Germany and Spain from the day of the first reported case (in every country) to 31 July of 2020. The daily report takes a moving window with the last 22, 12 and 19 days respectively. The last 5 days of historical data are not been taken in order to do error calculation and verify the model.

3.3.1. Italy

The total number of reported cases of Italy were taken from the Johns Hopkins University from January 31 to July 8 of 2020 and the daily forecast are from July 4 to July 8.

The parameters estimated for the forecast of figure 12 are: $\hat{\hat{\mathbf{f}}} = 213,702.42$, $\hat{\mathbf{q}} = 1.0995$ and $\hat{\mathbf{m}} = 0.33529$



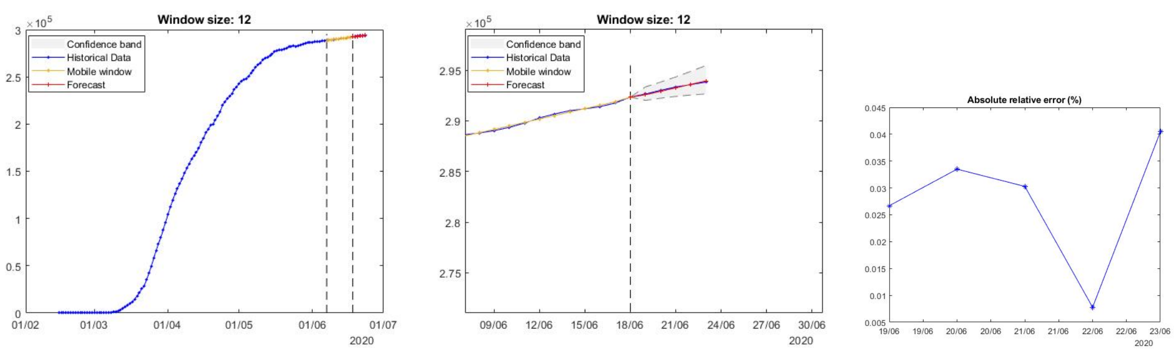
(a) (b) (c)

Figure 12: (a) Comparison between historical data of reported cases using a moving window of size 22 days and forecast of 5. (b) Detail of the forecast of 5 days and confidence band . (c) Absolute relative error of the forecasts.

3.3.2. Spain

The total number of reported cases of Italy were taken from Johns Hopkins University from February 1 to June 18 of 2020 and the daily forecast are from June 14 to June 18.

The parameters estimated for figure 13 are: $\hat{\hat{\mathbf{f}}} = 252,644.218$, $\hat{\mathbf{q}} = 0.93525$ and $\hat{\mathbf{m}} = 0.431685$



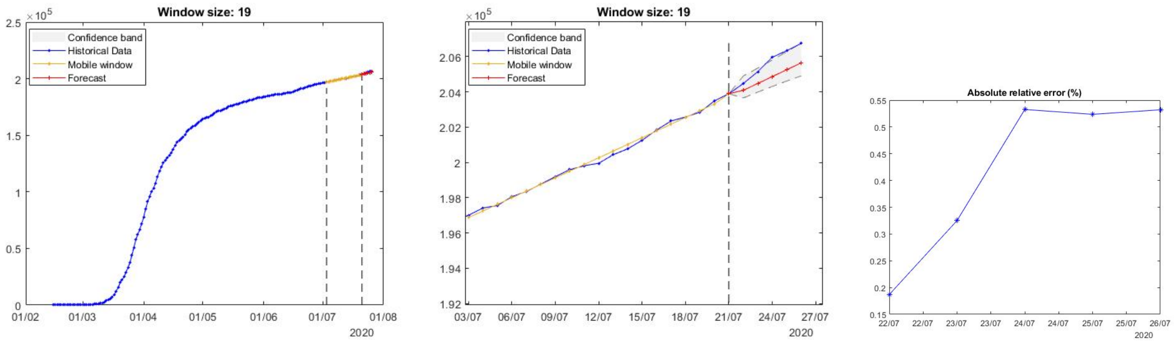
(a) (b) (c)

Figure 13: (a) Comparison between historical data of reported cases using a moving window of size 12 days and forecast of 5 days. (b) Detail of the forecast of 5 days and confidence band. (c) Absolute relative error of the forecasts.

3.3.3. Germany

The total number of reported cases of Italy were taken from Johns Hopkins University from January 28 to July 21 of 2020 and the daily forecast are from July 17 to July 21.

The parameters estimated for figure 14 are: $\hat{\hat{\mathbf{f}}} = 15,1695.72$, $\hat{\mathbf{q}} = 1.0399$ and $\hat{\mathbf{m}} = 0.68342$



(a) (b) (c)

Figure 14: (a) Comparison between historical data of reported case using a moving window of size 19 days and forecast of 5 days. (b) Detail of the forecast of 5 days and confidence band. (c) Absolute relative error of the forecasts.

4. Conclusions

In this research, a model based on q-exponential functions is proposed which has been inspired by the work developed in [20]. This model is very approximate or explains very well the dynamic behavior of the contagion curve since, for the 4 selected countries that were studied, an R^2 greater than 0.999 was obtained. The relative forecast mean error for all the countries is closed to zero and the standard deviation for Colombia, Mexico and Argentina is lower than 0.04 while the standard deviation for India is 0.15. For ten days forecast relative error between 0.1% and 5.4% were obtained, and moving windows forecast the lowest absolute relative error is around 0.006% and the highest is 0.54%.
When the curve of reported cases takes a concavity change in its behavior it can be modeled from the inflection point using the negative form of the q-exponential proposed model. Other important variables like recovered and deceased are likely to be explained towards the model proposed in this research. Although, this mathematical model and its parameter estimation are simple, offers good forecast and it counts with easy computational implementation. Also for a next-day forecast takes less than a second.

Acknowledgements

This research has been carried out in the framework of the project "Plataforma web para la recolección de datos, visualización, análisis, predicción y evaluación de estrategias de control de la enfermedad producida por SARS-CoV-2 mediante herramientas de modelación matemática, simulación e inteligencia artificial" which has been funded by the program MinCienciaTón (Covid-19 2020) of MinCinencias Colombia and EAFIT University through the agreement number 1216101576695.

References

[1] B. F. Maier, D. Brockmann, Effective containment explains subexponential growth in recent confirmed covid-19 cases in china, Science 368.

[2] J. Gao, P. Zheng, Y. Jia, H. Chen, Y. Mao, S. Chen, Y. Wang, H. Fu, J. Dai, Mental health problems and social media exposure during covid-19 outbreak, Plos one 15 (4) (2020) e0231924.

[3] M. Nicola, Z. Alsafi, C. Sohrabi, A. Kerwan, A. Al-Jabir, C. Iosifidis, M. Agha, R. Agha, The socio-economic implications of the coronavirus pandemic (covid-19): A review, International journal of surgery (London, England) 78 (2020) 185.

[4] D. Laborde, W. Martin, R. Vos, Poverty and food insecurity could grow dramatically as covid-19 spreads, Research Post, April 16 (2020) 2020.

[5] R. Dandekar, G. Barbastathis, Neural network aided quarantine control model estimation of global covid-19 spread, arXiv preprint: 2004.02752 1.

[6] R. V. P. Ashutosh Simha, S. Narayana, A simple stochastic sir model for covid-19 infection dynamics for karnataka after interventions – learning from european trends, arXiv preprint: 2003.11920 3.

[7] Y. Zhang, C. You, Z. Cai, J. Sun, W. Hu, X.-H. Zhou, Prediction of the covid-19 outbreak based on a realistic stochastic model, medRxiv.

[8] D. Fanelli, F. Piazza, Analysis and forecast of covid-19 spreading in china, italy and france, Chaos, Solitons & Fractals 134 (2020) 109761.

[9] O. S. Z. Liu, P. Magal, G. Webb, Predicting the cumulative number of cases for the covid-19 epidemic in china from early data, arXiv preprint: 2002.12298 1.

[10] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday, et al., Early dynamics of transmission and control of covid-19: a mathematical modelling study, The lancet infectious diseases.

[11] X. Chen, B. Yu, First two months of the 2019 coronavirus disease (covid-19) epidemic in china: real-time surveillance and evaluation with a second derivative model, Global health research and policy 5 (1) (2020) 1–9.

[12] G. L. Vasconcelos, A. M. Macêdo, R. Ospina, F. A. Almeida, G. C. Duarte-Filho, A. A. Brum, I. C. Souza, Modelling fatality curves of covid-19 and the effectiveness of intervention strategies, PeerJ 8 (2020) e9421.

[13] R. A. R. J. P. K.Roosaa, Y.Leea, G.Chowella, Real-time forecasts of the covid-19 epidemic in china from february 5th to february 24th, 2020, ScienceDirect preprint 5 (2020) 256–263.

[14] R. Cherniha, V. Davydovych, A mathematical model for the coronavirus covid-19 outbreak, arXiv preprint arXiv:2004.01487.

[15] F. Petropoulos, S. Makridakis, Forecasting the novel coronavirus covid-19, PloS one 15 (3) (2020) e0231236.

[16] O. Torrealba-Rodriguez, R. Conde-Gutiérrez, A. Hernández-Javier, Modeling and prediction of covid-19 in mexico applying mathematical and computational models, Chaos, Solitons & Fractals (2020) 109946.

[17] I. N. de Salud, Coronavirus (covid-19) en Colombia, `https://www.ins.gov.co/Noticias/Paginas/Coronavirus.aspx` (2020).

[18] U. o. M. Johns Hopkins, Coronavirus resource center, `https://coronavirus.jhu.edu/data/new-cases` (2020).

[19] D. McAnally, q-exponential and q-gamma functions. i. q-exponential functionsa, Journal of Mathematical Physics 36 (1) (1995) 546–573.

[20] C. Tsallis, What are the numbers that experiments provide, Quimica Nova 17 (6) (1994) 468–471.

[21] C. Tsallis, I. nonextensive statistical mechanics and thermodynamics: Historical background and present status, in: Nonextensive statistical mechanics and its applications, Springer, 2001, pp. 3–98.