

PAPER • OPEN ACCESS

## Computational estimation from a statistical physics approach and its contributions to the Covid-19 in Colombia

To cite this article: L Ramirez-Carvajal *et al* 2020 *J. Phys.: Conf. Ser.* **1702** 012016

View the [article online](#) for updates and enhancements.

You may also like

- [Construction of public policy of higher education with participation of the academic community. Colombian case for recognition of virtual and distance learning programs](#)  
F Cantor
- [Geographic overlaps between priority areas for forest carbon-storage efforts and those for delivering peacebuilding programs: implications for policy design](#)  
Augusto Castro-Nunez, Ole Mertz and Chrystian C Sosa
- [Smart Grids and their Applicability for the Development of the Electricity Sector for Colombia in the year 2050](#)  
J Viola and C Aceros



The Electrochemical Society  
Advancing solid state & electrochemical science & technology

### 241st ECS Meeting

May 29 – June 2, 2022 Vancouver • BC • Canada

Abstract submission deadline: Dec 3, 2021

Connect. Engage. Champion. Empower. Accelerate.  
**We move science forward**



**Submit your abstract**



# Computational estimation from a statistical physics approach and its contributions to the Covid-19 in Colombia

L Ramírez-Carvajal<sup>1</sup>, K Puerto-López<sup>1</sup>, and B Medina-Delgado<sup>1</sup>

<sup>1</sup> Universidad Francisco de Paula Santander, San José de Cúcuta, Colombia

E-mail: luiseduardorc@ufps.edu.co

**Abstract.** The objective of this article is to present a computational estimate from a statistical physics approach and its contributions to Covid-19 in Colombia. Based on the daily data of contagions, recoveries and deaths, during the months of March to July, the estimation of the behavior of the epidemic was made using the nonlinear regression method with adjustment of curves by minimum squares. Highlighting the benefits that this method presents in the study of physical phenomena, it was used in the present research developing two types of modeling: exponential and Gaussian, and with these some predictions were made. The coefficients of determination of the exponential model were: 0.9641 for contagions, 0.9400 for recoveries and 0.9788 for deaths, and those of the Gaussian model were: 0.9799 for contagions, 0.9606 for recoveries and 0.9894 for deaths, showing a good correlation between the models and the real behavior of the pandemic, being the Gaussian one, the most approximate. This was also evidenced by comparing the prognosis of both models with the actual data for the first 13 days of August, concluding that the pandemic is beginning to mitigate, and the curve is flattening out.

## 1. Introduction

Covid-19 is a virus unknown to mankind. Its understanding is based on the hard work of countless researchers [1]. Many of them have analyzed the epidemic from a statistical point of view using different models, such as susceptible, exposed, infected and recovered (SEIR) [2] and Susceptible, Infected and recovered (SIR) [3], specialized in the prediction of infectious diseases, however, these could generate imprecise results due to the systematic variations in the prognosis curve and the complexity of the epidemic [2], therefore, it is extremely important for science and humanity to have more research using different methods such as: non-linear regression modeling [4].

Non-linear regression is a method that uses least-squares curve fitting [5]. This method emerged from the fields of astronomy and geodesy. The first scientists to contribute to this method were Carl Friedrich Gauss, Adrien-Marie Legendre and Robert Adrain in the XVIII century [6]. It has been widely used in different areas of knowledge such as statistical mechanics, a discipline that was born in the nineteenth century with the contributions of Rudolf Clausius, James Clerk Maxwell and especially Ludwig Boltzmann [7]. All these scientists established the basis of statistical physics and their contributions still have great importance in research of recent years, such as those carried out by Flórez and Laguado, for example, in computational fluid dynamics [8], that of Plaza in the modeling of physical and natural phenomena [9] or Vera, Delgado and Sepulveda, in solar energy [10].

Statistical mechanics is as a link between a macroscopic world treated as continuous and a microscopic world of discrete nature [6]; as the Covid-19, whose discrete components are the data reported each day, and these describe behavior that can be modeled continuously over time [11], this, in



a computational way, with the objective of optimizing the estimation and decision making [12].

## 2. Materials and methods

The research was carried out using a descriptive and applied methodology that can be seen in Figure 1.

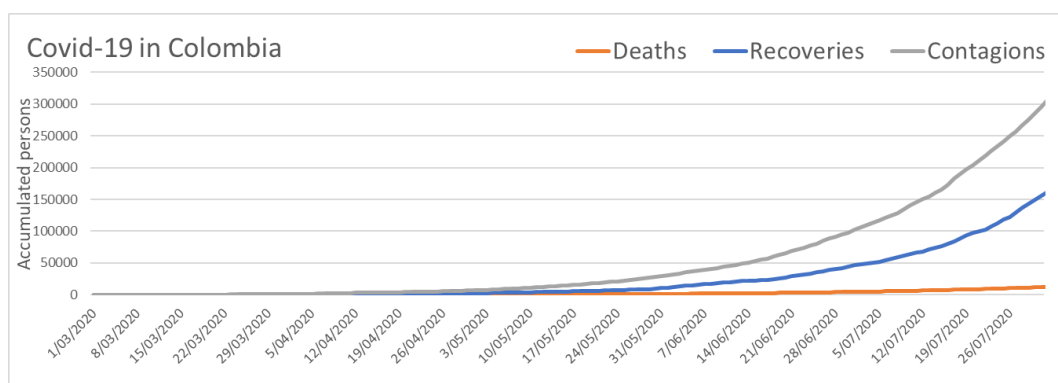


**Figure 1.** Methodology used in the project.

### 2.1. Collecting data

Recent studies such as those conducted by Diaz Pinzón [13], Verbel, Mejía, Manjarres and Troncoso [14], show that the pandemic does not have a defined behavior. Figure 2 for example, shows the behavior of the pandemic in Colombia. This figure was made with data on contagions, recoveries and deaths reported by the "Instituto Nacional de Salud" (INS) [15].

Figure 2 shows that the Colombia curve has a growing exponential behavior. Although the effects of the pandemic were initially mitigated, the curve has been steadily increasing, reaching levels of contagion and deaths similar to those countries that initially did not take preventive measures and were greatly impacted by the pandemic. In those countries, drastic measures were taken due to the high rate of contagions and deaths and a flattening of the curve was witnessed. In Colombia, on the contrary, measures are being taken, but the curve continues to increase exponentially.



**Figure 2.** Covid-19 in Colombia.

### 2.2. Modelling

Using non-linear regression with a technological tool, two types of mathematical models were chosen: exponential [16] and Gaussian [5]. Exponential due to the behavior of the curve in Colombia, which does not yet show any kind of flattening. And the Gaussian one to estimate a flattening in the following months.

### 2.3. Model analysis

The number of terms indicated were chosen to obtain a better result in the behavior of the curve and in the calculation of the determination coefficient.

### 2.4. Prediction development

With the exponential model, predictions were made for the months of August and September, considering that the curve is close to the flattening point, and with the Gaussian model, predictions were made for the months of August, September, October, November and December. With both models the error between the predicted and actual data was calculated for the first 13 days of August.

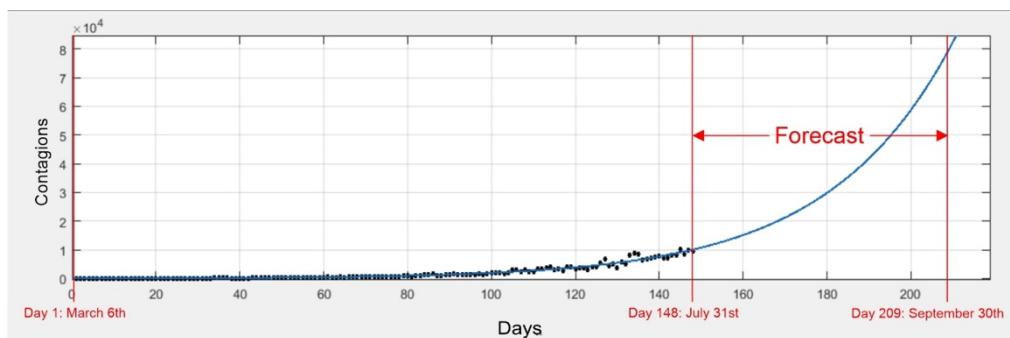
### 3. Results and discussion

Below are the two types of models estimated from data collected between March 6, 2020, the day the first contagion occurred, and July 31, 2020.

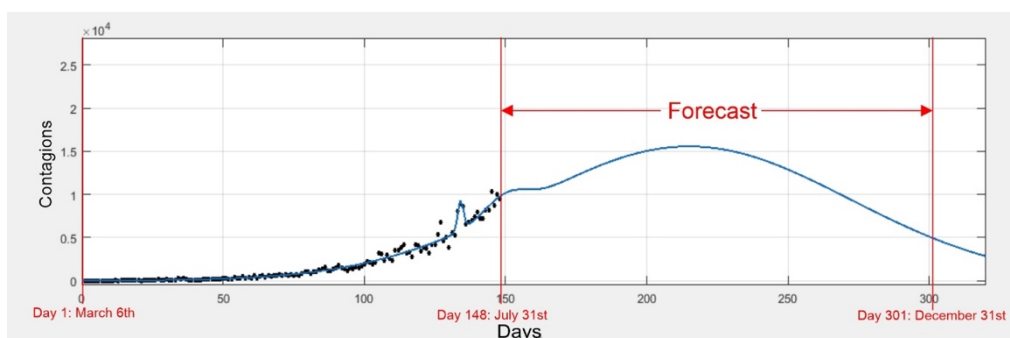
#### 3.1. Models

Figure 3 shows the exponential model for the contagion curve and Figure 4 shows the Gaussian model. The same modeling was done with the recoveries and deceased data. The exponential model in Figure 3 is appropriate to predict a possible increase in contagions until day 209, which is equivalent to September 30th, 2020. The same modeling was used for the recoveries curve and the deaths curve.

The Gaussian model predicts a possible flattening of the curve between the days 210 and 270, which is equivalent to the months of August and September 2020. The same modeling was used for the recoveries curve and the deaths curve. Two models were used for the forecasts: The exponential model predicts an increase in the number of contagions per day, a phenomenon that was evident in the behavior of the real data taken, and, if it is not contained, it can produce more than 80000 contagions per day after September 30th, 2020. On the contrary, the Gaussian model predicts a flattening of the curve, whose peak does not reach 20000 contagions per day, and by December 31, 2020, it expects to be below 300 contagions per day.



**Figure 3.** Exponential model for contagions.



**Figure 4.** Gaussian model for contagions.

#### 3.2. Coefficients of determination

Table 1 shows the determination coefficients of each of the models. The determination coefficients showed a good correlation between the models and the real behavior of the pandemic, being more approximate the Gaussian model.

**Table 1.** Coefficients of determination.

Model	$R^2$	Model	$R^2$
Contagions per day, exponential	0.9641	Contagions per day, Gaussian	0.9799
Recoveries per day, exponential	0.9400	Recoveries per day, Gaussian	0.9606
Deaths per day, exponential	0.9788	Deceased per day, Gaussian	0.9894

### 3.3. Prediction development

Table 2 shows the predictions of the exponential model for some days in August and September and Table 3 shows the predictions of the Gaussian model for a possible flattening of the curve. These predictions presented refer to accumulated data per day. With the exponential and Gaussian model, we calculated the accumulated contagions per day, the accumulated recoveries per day and the accumulated deaths per day and these were the data that were compared with the real ones.

**Table 2.** Exponential model predictions.

Date	Total contagion	Total recoveries	Total, deaths	Date	Total contagion	Total recoveries	Total death
Aug 1	311690	164519	12105	Sep 01	894255	996090	29685
Aug 15	501981	333738	18169	Sep 15	1437855	3072007	44427
Aug 31	864413	926131	28842	Sep 30	2390798	12191459	68384

**Table 3.** Gaussian model predictions.

Date	Total, contagion	Total, recoveries	Total, deaths	Date	Total, contagion	Total, recoveries	Total, deaths
Aug 1	305131	159820	11568	Oct 15	1305776	400487	42325
Aug 15	452551	213615	16465	Nov 1	1556451	422088	47927
Sept 1	652638	277827	23478	Nov 15	1739235	431755	51414
Sept 15	846326	326640	29714	Dec 1	1912137	437190	54203
Sept 30	1073203	369545	36318	Dec 15	2029065	439233	55770
Oct 1	1088705	371987	36743	Dec 31	2125513	440132	56829

It is observed that after a possible flattening of the Gaussian curve there will be fewer contagions and deaths by December 31st, 2020 than predicted by the exponential model for September 30th, 2020. Table 4 shows the comparison between actual contagions and models for the first 13 days of August. Table 4 shows that the errors of the Gaussian model were all below 1, while those of the exponential model were all above 1.5. Table 5 shows the comparison between actual recoveries and models for the first 13 days of August. In this table we can see that the errors started to increase above 5% for both models, after August 4th, 2020. What allows us to see this phenomenon is that the curve of recoveries per day does not seem to follow a defined pattern but remains between an exponential and a Gaussian behavior. For example, for the 13 of August of the 2020, the exponential model predicted almost 300000 recoveries, whereas the Gaussian model just over 200000, and recoveries were 250971, that is to say, almost to half of a model and another one.

**Table 4.** Models vs. real data for contagions, August 1 - August 13, 2020.

Day	Date	Contagions actual	Exponential model	Error% in exponential model	Gaussian model	Error% in Gaussian model
149	Aug 1	306181	311690	1.799	305131	0.342
150	Aug 2	317651	322496	1.525	315325	0.732
151	Aug 3	327850	333674	1.776	325663	0.667
152	Aug 4	334979	345238	3.062	336106	0.336
153	Aug 5	345714	357199	3.322	346622	0.262
154	Aug 6	357710	369573	3.316	357180	0.148
155	Aug 7	367196	382372	4.132	367757	0.152
156	Aug 8	376870	395612	4.973	378339	0.389
157	Aug 9	387481	409309	5.633	388915	0.370
158	Aug 10	397623	423477	6.502	399485	0.468
159	Aug 11	410453	438133	6.743	410054	0.097
160	Aug 12	422519	453293	7.283	420630	0.447
161	Aug 13	433805	468976	8.107	431228	0.594

**Table 5.** Models vs. real data for recoveries, August 1 - August 13, 2020.

Day	Date	Recoveries actual	Exponential model	Error% in exponential model	Gaussian model	Error% in Gaussian model
149	Aug 1	162400	164519	1.287	159820	1.6143
150	Aug 2	168876	172324	2.000	164623	2.5834
151	Aug 3	175420	180596	2.866	168971	3.8166
152	Aug 4	182058	189371	3.861	172987	5.2437
153	Aug 5	187984	198689	5.387	176792	6.3306
154	Aug 6	193876	208593	7.055	180481	7.4218
155	Aug 7	199861	219132	8.794	184120	8.5493
156	Aug 8	205758	230356	10.678	187747	9.5932
157	Aug 9	213712	242325	11.807	191384	11.666
158	Aug 10	221901	255099	13.013	195039	13.772
159	Aug 11	230815	268750	14.115	198715	16.153
160	Aug 12	240275	283351	15.202	202412	18.705
161	Aug 13	250971	298987	16.059	206129	21.754

Table 6 shows the comparison between actual deaths and models for the first 13 days of August. This table shows that the deaths curve followed the Gaussian model, since the errors were greater in the exponential model. Table 7 shows the average errors between the models and the real data for contagions and deaths during the first 13 days of August. It is observed that the mean error between the real data and the forecast is lower for the Gaussian model in the contagions and deaths, and lower in the recoveries for the exponential model.

**Table 6.** Models vs. real data for deaths, August 1 - August 13, 2020.

Day	Date	Deaths actual	Exponential model	Error% in exponential model	Gaussian model	Error% in Gaussian model
149	Aug 1	11616	12105	4.039	11568	0.414
150	Aug 2	11871	12463	4.750	11886	0.126
151	Aug 3	12167	12830	5.167	12209	0.344
152	Aug 4	12427	13208	5.913	12537	0.877
153	Aug 5	12692	13597	6.655	12870	1.383
154	Aug 6	12941	13998	7.551	13208	2.021
155	Aug 7	13214	14410	8.299	13551	2.486
156	Aug 8	13501	14834	8.986	13899	2.863
157	Aug 9	13729	15270	10.091	14252	3.669
158	Aug 10	13951	15720	11.253	14610	4.510
159	Aug 11	14151	16182	12.550	14972	5.483
160	Aug 12	14314	16657	14.066	15338	6.676
161	Aug 13	14476	17147	15.577	15710	7.854

Table 7 shows the average errors between the models and the real data for contagions and deaths during the first 13 days of August. It is observed that the mean error between the real data and the forecast is lower for the Gaussian model in the contagions and deaths, and lower in the recoveries for the exponential model.

**Table 7.** Average errors in models vs. actual data 1 - 13 August 2020.

	Average error
Real vs. exponential modeling	4.475
Real deaths vs. exponential model	8.838
Real recoveries vs. exponential model	8.625
Real-life accounts vs. Gaussian model	0.385
Actual deaths vs. Gaussian model	2.977
Real-life recoveries vs. Gaussian model	9.785

#### 4. Conclusions

Approaching the research from a statistical physics approach was successful, because the behavior of covid-19 can be modeled as a macro system, continuous in time that depends on discrete elements, reported day by day. The use of technological instruments in the statistical analysis and the prediction methods facilitates the models' estimation, making the process more dynamic and shortening the necessary time to reach conclusions that can benefit the decision making. Differences that may arise between prognostic curves and actual data relate to the nation's response to the pandemic, the presence of asymptomatic patients, prevention measures, data collection capacity, processing times, and follow-up of infected patients, but the results presented here are nonetheless relevant to the analysis of the pandemic. The determination coefficients showed a good correlation between the models and the pandemic, with the Gaussian model being the most accurate, allowing us to infer that the pandemic will begin to show mitigation from day 200, which is equivalent to the second week of September 2020.

The behavior of the pandemic depends on the actions taken by the authorities to mitigate the contagion and the compliance of citizens to these actions, since the exponential curve shows that, if the current growth is not mitigated, we could have 2390798 contagions and more than 68000 deaths by September 30th, 2020. If the pandemic responds to Gaussian behavior it is expected to have more than 2 million contagions and about 56800 deaths by December 31st, 2020, which is a much lower number than expected by the exponential curve, however, are still worrying figures that should lead the country to strengthen prevention measures.

#### References

- [1] Dong Y, Guo W, Ren Q 2020 *Journal of Physics: Conference Series* **1575** 012228:1
- [2] Ortega Lenis D, Arango Londoño D, Muñoz E, Cuartas E, Caicedo D, Mena J et al. 2020 *Revista Salud Pública* **22(2)** 1
- [3] Manrique Abril F G, Agudelo Calderón A, Gonzales Chordá V M, Gutiérrez Lesmes O, Téllez Piñeres F, Herrera Amaya G 2020 *Revista Salud Pública* **22(2)** 1
- [4] Obando Bastidas J A, Peña Pita A P, Obando Vargas L N, Franco Montenegro A 2020 *Revista Habanera de Ciencias Médicas* **19(1)** 1
- [5] Chapra S C, Canale R P 2015 *Métodos Numéricos para Ingenieros* (México: McGraw-Hill)
- [6] Lemagne Pérez J, Calzadilla Méndez A 2012 *Revista Pensamiento Matemático* **2** 1
- [7] Guzman R, Cervera J A 2006 *Llull Revista de la Sociedad Española de Historia de las Ciencias y de las Técnicas* **29(64)** 331
- [8] Florez Serrano E G, Laguado Ramírez I 2011 *Ingeniería y Universidad* **15(2)** 469
- [9] Plaza Gálvez L F 2016 *IE Revista de Investigación Educativa de la REDIECH* **7(13)** 47
- [10] Vera Dávila Á G, Delgado Ariza C, Sepulveda Mora S B 2018 *Revista de Investigación, Desarrollo e Innovación* **8(2)** 343
- [11] Singer H M 2020 *Physical Biology* **17** 055001:1
- [12] Minnaard C 2010 *Modos de Regresión Lineales y no Lineales: Su Aplicación en Problemas de Ingeniería* (Argentina: Universidad Nacional de Lomas de Zamora)
- [13] Diaz Pinzón J E 2020 *Repertorio de Medicina y Cirujía* **29(1)** 65
- [14] Verbel Buelvas L, Mejía Villareal J, Manjarres Gomez H, Troncoso Palacio A 2020 *BILO* **2(1)** 38
- [15] Instituto Nacional de Salud (INS) 2020 *Coronavirus (COVID - 19) en Colombia* (Bogotá: Instituto Nacional de Salud)
- [16] Purcell E, Varberg D, Rigdon S 2007 *Cálculo Diferencial e Integral* (México: Pearson Educación)