

CIENCIAS TECNOLÓGICAS  
ARTÍCULO ORIGINAL**Ajuste de curvas de crecimiento poblacional aplicadas a la COVID-19 en Cuba****Adjustment of population growth curve applied to COVID-19 in Cuba**Juan Felipe Medina Mendieta<sup>1</sup>  , Manuel E. Cortés Cortés<sup>1</sup> , Manuel Cortés Iglesias<sup>1</sup> <sup>1</sup> Universidad de Cienfuegos “Carlos Rafael Rodríguez”. Cienfuegos, Cuba.**Cómo citar este artículo**

Medina Mendieta JF, Cortés Cortés ME, Cortés Iglesias M. Ajuste de curvas de crecimiento poblacional aplicadas a la COVID-19 en Cuba. Rev haban cienc méd [Internet]. 2020 [citado ]; 19(Supl.):e3353. Disponible en: <http://www.revhabanera.sld.cu/index.php/rhab/article/view/3353>

Recibido: 05 de mayo del 2020.

Aprobado: 14 de mayo del 2020.

**RESUMEN**

**Introducción:** Cuba ha sido afectada por la COVID-19. Todas las provincias del país han presentado casos confirmados de la enfermedad. Se han llevado a cabo medidas por parte del gobierno y el sistema de salud, para contrarrestar el contagio de persona a persona. Es de gran ayuda contar con estimaciones de casos confirmados para las decisiones.

**Objetivos:** Obtener predicciones para los picos de

casos confirmados y cantidad total de estos para algunas provincias de Cuba y para todo el país.

**Material y Métodos:** Estudio de tipo predictivo de curvas de crecimiento poblacional. Se analizan los datos correspondientes a los primeros 52 días de afectación de la enfermedad en el país para estimar los modelos y aplicar el método de los mínimos cuadrados para modelos no lineales con respecto a los parámetros. Se utilizan el



coeficiente de determinación ajustado, el criterio de información de Akaike y el error estándar de los residuos para medir la bondad del ajuste de los modelos. Se estudian las provincias del país que presentan una tasa de infectados por cien mil habitantes mayor que 14,71 y el país en su conjunto.

**Resultados:** La bondad de ajuste de los modelos utilizados en las localidades estudiadas y en el

país es alta, lo cual permite su confiabilidad para los pronósticos efectuados.

**Conclusiones:** Las predicciones plantean que las cinco localidades analizadas presentan su pico de contagio en abril al igual que Cuba.

**Palabras claves:** COVID-19, estudio predictivo, curvas de crecimiento poblacional, Cuba.

### ABSTRACT

**Introduction:** Cuba and all its provinces have been affected by COVID-19 disease. The government and the health system have taken measures to avoid contagion from person to person. To take these measures it is important to have estimates of the rate of infection.

**Objectives:** To obtain predictions for the peak of infected cases and the total number for some Cuban provinces and the whole country.

**Material and Methods:** Predictive study of population growth curves. Data from the first 52 days of the disease in the country are processed to estimate the models and to apply the method of least squares estimation of nonlinear parameters. The adjusted coefficient of

determination, the Akaike information criterion and the standard error of the residuals are used to measure the goodness of fit of the models. The provinces that present a rate of infection per 100,000 inhabitants greater than 14,71 and the country as a whole are studied.

**Results:** The goodness of fit of the models used in the provinces studied and the country is high, which allows them to be reliable for predictions.

**Conclusions:** The predictions suggest that the five provinces analyzed and Cuba show their peak of contagion in April.

**Keywords:** COVID-19, predictive study, population growth curves, Cuba.

### INTRODUCCIÓN

El nuevo coronavirus Sars-CoV-2, causante de la enfermedad COVID-19, presenta una rápida propagación y alto porcentaje de mortalidad, comparado con otros coronavirus.

Cuba se encuentra afectada por la COVID-19 y ha adoptado medidas, con el fin de contrarrestar su propagación en la población. Sin embargo, como ha sucedido en otros países afectados por esta

enfermedad, se ha manifestado un crecimiento aproximadamente exponencial de casos confirmados en el país; las provincias de occidente presentan mayor número de casos que las de oriente. Entre las provincias con mayor número de casos confirmados destaca La Habana. Este comportamiento es lógico debido a que en ella se concentra aproximadamente la



quinta parte de la población nacional y presenta un mayor movimiento interno y externo.

La dirección del país ha llevado a cabo medidas de gran magnitud, en coordinación con el Ministerio de Salud Pública, la Defensa Civil y otras entidades. Se han realizado pronósticos de la cantidad de personas que podrán estar en centros de confinamiento, hospitalizadas o con necesidad de cuidados intensivos, así como de personal médico disponible, entre muchas otras cuestiones. Los pronósticos de la cantidad total de personas contagiadas por provincia y el o los días de mayor número de contagios se hace imprescindible en este contexto y para realizarlo se utilizan distintas modelaciones matemáticas.

Una modelación clásica aplicada a las epidemias consiste en los modelos SIR de ecuaciones diferenciales, cuyo nombre proviene de: S (población susceptible), I (población infectada) y R (población recuperada o removida del sistema). Esta modelación tiene presente una serie de factores con el fin de expresar lo mejor posible las realidades de las epidemias y ha sido utilizada con éxitos en la epidemia provocada por la COVID-19, en algunas regiones.<sup>(1,2)</sup> Sin embargo, en muchas ocasiones, presenta gran complejidad en la modelación debido a los diferentes factores a tener presente, alto nivel de procesamiento para la obtención de los modelos y alta complejidad en la interpretación de las soluciones obtenidas.

Una modelación menos compleja está basada en ajuste de modelos predictivos de crecimiento

poblacional del tipo logístico y exponencial. Existen diferentes modelos que describen el crecimiento de poblaciones.<sup>(3)</sup> Estos pueden ser aplicados al incremento de la cantidad de personas infectadas por una enfermedad de contagio. Entre estos modelos destacan aquellos de crecimiento logístico, de Gompertz, de Weibull. Estos han sido utilizados con éxito en la epidemia de la COVID-19 para los casos de China,<sup>(4,5,6)</sup> Italia,<sup>(7,8,9)</sup> Corea del Sur, Costa Rica<sup>(10)</sup> y varios países.<sup>(11)</sup>

El objetivo de utilizar esta modelación matemática, consiste en obtener los parámetros de los modelos que mejor se ajusten a los datos reales, para utilizar estos en predicciones futuras. Para la obtención de los parámetros, se utiliza el método de los mínimos cuadrados para modelos no lineales con respecto a los parámetros (MCNL).

Es importante seleccionar el o aquellos modelos que mejor describan la situación en estudio y presenten validez para realizar pronósticos. Debido a esto se pueden formular las siguientes interrogantes: ¿cuál o cuáles de los modelos de crecimiento poblacional aplicados se ajustan mejor a los datos reales? ¿serán adecuados el o los modelos seleccionados para realizar pronósticos futuros?

El **objetivo** de la presente investigación es obtener predicciones para los picos de casos confirmados y cantidad total de estos para algunas provincias de Cuba y para todo el país.

## MATERIAL Y MÉTODOS

El estudio predictivo realizado presenta modelos de crecimiento poblacional con adecuaciones.

Fueron utilizados 7 modelos entre variaciones del modelo logístico y los modelos de crecimiento



Gompertz y de Weibull. En estos, la variable dependiente está representada por los casos acumulados confirmados a la COVID-19 diariamente y la independiente por el tiempo transcurrido, luego de haberse registrado los primeros casos. La cantidad de parámetros a

obtener varía entre 4 y 7 en dependencia del modelo aplicado y ha sido utilizado el método de los MCNL para su obtención. El valor de la significación a priori alfa escogido es 0,05. Los modelos utilizados en la presente investigación se exponen a continuación:

- M1:  $f(t) = c + \frac{d - c}{1 + e^{b \cdot (\ln t - \ln a)}}$ . Modelo logarítmico-logístico de cuatro parámetros.<sup>(12,13)</sup>
- M2:  $f(t) = c + \frac{d - c}{\left(1 + e^{b \cdot (\ln t - \ln a)}\right)^h}$ . Modelo logarítmico-logístico de cinco parámetros.<sup>(12,14)</sup>
- M3:  $f(t) = c + (d - c) \cdot e^{-e^{b \cdot (t - a)}}$ . Modelo de crecimiento de Gompertz de cuatro parámetros.<sup>(12,15)</sup>
- M4:  $f(t) = c + (d - c) \cdot \left(1 - e^{-e^{b \cdot (\ln t - \ln a)}}\right)$ . Modelo de crecimiento de Weibull de cuatro parámetros.<sup>(12,16)</sup>
- M5:  $f(t) = c + \frac{d - c + h \cdot t}{1 + e^{b \cdot (\ln t - \ln a)}}$ . Modelo logístico de Brain and Cousens de cinco parámetros.<sup>(17)</sup>
- M6:  $f(t) = c + \frac{d_1 - c}{1 + e^{b_1 \cdot (\ln t - \ln a_1)}} + \frac{d_2}{1 + e^{b_2 \cdot (\ln t - \ln a_2)}}$ . Modelo de dos fases de dosis-respuesta de siete parámetros.<sup>(18)</sup>
- M7:  $f(t) = c + \frac{d - c}{1 + g(t) \cdot e^{b_1 \cdot (\ln t - \ln a)} + (1 - g(t)) \cdot e^{b_2 \cdot (\ln t - \ln a)}}$ ,  
 $g(t) = \frac{1}{1 + e^{\left(\frac{2b_1 \cdot b_2}{|b_1 + b_2|} \cdot (\ln t - \ln a)\right)}}$ . Variación del modelo logístico de cinco parámetros.<sup>(19)</sup>

El estudio fue realizado desde mediados de marzo hasta finales de abril, en medio de la afectación de la COVID-19 en Cuba, por un grupo

de profesores del Departamento de Matemáticas de la Universidad de Cienfuegos “Carlos Rafael Rodríguez”, en colaboración con el



Departamento de Tecnología Educativa de este Centro.

Se estudiaron los datos reportados por Cuba referente a la cantidad de casos confirmados a la COVID-19; se comenzó el registro el 11 de marzo, fecha en la que se reportaron los primeros casos positivos de la enfermedad en el país. Los datos fueron estudiados hasta el 2 de mayo de 2020. Estos datos fueron tomados del sitio [covid19cubadata.github.io](https://covid19cubadata.github.io), los cuales son actualizados diariamente, por colaboradores del sitio *web*, y compilados en un archivo de datos de extensión csv, a partir de los partes oficiales que, en conferencia de prensa, se ofrecen a todo el

país.

Para el estudio, y hasta la fecha de escrito este informe de investigación (3 de mayo de 2020), se seleccionaron los 5 territorios más afectadas por la COVID-19, en función de su cantidad de habitantes. Los territorios seleccionados para el estudio fueron aquellos con tasa de infectados por cien mil habitantes mayor que 14,71, debido a que el país, en esta fecha presentaba este índice. En la Tabla 1 se muestra, por provincia, la cantidad de casos confirmados y las respectivas tasas.

**Tabla 1.** Cantidad de casos confirmados a la COVID-19 y tasa por cada cien mil habitantes, por cada provincia y el Municipio Especial en Cuba

Provincia/ Municipio Especial	COVID-19 en Cuba (03/05/2020)	
	Casos positivos	Tasa/10 <sup>5</sup> habitantes
Isla de la Juventud	41	48,93
La Habana	776	36,41
Villa Clara	192	24,59
Ciego de Ávila	84	19,31
Matanzas	124	17,35
Sancti Spiritus	64	13,74
Mayabeque	46	12,01
Holguín	89	8,66
Pinar del Río	50	8,50
Artemisa	31	6,07
Cienfuegos	24	5,90
Camagüey	37	4,82
Santiago de Cuba	48	4,57
Guantánamo	17	3,34
Las Tunas	14	2,62
Granma	12	1,46



Los territorios seleccionados para el estudio fueron: Isla de la Juventud (48,93), La Habana (36,41), Villa Clara (24,59), Ciego de Ávila (19,31) y Matanzas (17,35). En la Isla de la Juventud se registraron los primeros casos el 04/04/2020; en La Habana, el 16/03/2020; en Villa Clara, el 13/03/2020; en Ciego de Ávila, el 23/03/2020 y en Matanzas, el 20/03/2020.

Las variables cuantitativas discretas, analizadas en la investigación son:

Variable independiente: Cantidad de días transcurridos desde que se confirmaron el o los primeros casos positivos a la COVID-19 por localidad y en el país.

Variable dependiente: Cantidad de casos acumulados diarios, confirmados a la COVID-19 por localidad o en el país.

Para el tratamiento de los datos y el procesamiento computacional se utilizaron *scripts* programados en el lenguaje de programación R 3.6.1, de procesamiento numérico.<sup>(20)</sup> Los *scripts* pueden ser descargados desde la dirección <https://clasesvirtuales.ucf.edu/cu/course/view.php?id=1147>

Para el procesamiento de los modelos utilizados y la implementación del método de los MCNL se empleó la instrucción *drm* del paquete *drc* de R. Este paquete presenta herramientas para el análisis y el ajuste de modelos no lineales con respecto a sus parámetros, conocidos como modelos de dosis-respuesta, utilizados en Ciencias Médicas, Biológicas, Agropecuarias, Técnicas y Sociales. Presenta implementación de algoritmos eficientes para la obtención de los

parámetros, pruebas de hipótesis para estudiar la bondad de ajuste de estos modelos y el contraste de criterios para selección del mejor o mejores modelos que se ajustan a una serie de datos. Mayor documentación puede ser encontrada en <http://www.bioassay.dk>.

Para el cálculo de los puntos de inflexión en los modelos M5, M6 y M7 se utilizaron métodos numéricos mediante la instrucción *uniroot* del paquete *stat* de R.

Para la selección del modelo con mejor ajuste se contrastaron varios criterios:

- Coeficiente de determinación ajustado ( $R^2$ ), adecuado para modelos no lineales. Evalúa la efectividad que presenta la variable independiente para medir la variable dependiente.<sup>(21)</sup>
- Criterio de información de Akaike (AIC). Expone una medida de la calidad relativa de un modelo al contrastar la bondad del ajuste con la complejidad del mismo.<sup>(22)</sup>
- Error estándar de los residuos (RSE). Presenta una desviación media ponderada del modelo con respecto a los valores experimentales.<sup>(23)</sup>

Para diferentes modelos, valores del coeficiente de determinación ajustados cercanos a 1 y menores valores de AIC y de errores estándares de los residuos, ponen de manifiesto modelos con mayor adecuación.

Para verificar la bondad del ajuste, se aplicó la prueba de Neill, mediante la instrucción *neill.test* del paquete *drc* de R, adecuada para modelos no lineales con respecto a los parámetros y que utiliza técnicas de agrupamiento en caso de no contar con réplicas.<sup>(24)</sup>



**RESULTADOS**

*Estudio de caso de las provincias y el municipio especial Isla de la Juventud*

En la Tabla 2 se presentan los resultados del

procesamiento realizado para cada uno de los 5 territorios afectados. Se muestran los criterios de ajuste utilizados en la investigación.

**Tabla 2** - Criterios de ajuste para cada modelo procesado para los datos de cada territorio COVID-19

Modelos (02/05/2020)	Provincia/Municipio especial				
	Isla de la Juventud	La Habana	Villa Clara	Ciego de Ávila	Matanzas
<b>M1</b>	R <sup>2</sup> =0,974 AIC=139 RSE=2,39	R <sup>2</sup> =0,998 AIC=361 RSE=9,92	R <sup>2</sup> =0,987 AIC=375 RSE=9,06	R <sup>2</sup> =0,988 AIC=223 RSE=3,40	R <sup>2</sup> =0,989 AIC=254 RSE=4,05
<b>M2</b>	R <sup>2</sup> =0,983 AIC=127 RSE=1,95	R <sup>2</sup> =0,999 AIC=359 RSE=9,53	R <sup>2</sup> =0,995 AIC=331 RSE=5,81	R <sup>2</sup> =0,989 AIC=220 RSE=3,28	R <sup>2</sup> =0,989 AIC=255 RSE=4,06
<b>M3</b>	R <sup>2</sup> =0,975 AIC=137 RSE=2,31	R <sup>2</sup> =0,999 AIC=358 RSE=9,47	R <sup>2</sup> =0,983 AIC=387 RSE=10,2	R <sup>2</sup> =0,988 AIC=223 RSE=3,41	R <sup>2</sup> =0,991 AIC=246 RSE=3,72
<b>M4</b>	R <sup>2</sup> =0,977 AIC=135 RSE=2,25	R <sup>2</sup> =0,998 AIC=362 RSE=9,84	R <sup>2</sup> =0,991 AIC=356 RSE=7,49	R <sup>2</sup> =0,988 AIC=222 RSE=3,40	R <sup>2</sup> =0,989 AIC=253 RSE=4,01
<b>M5</b>	R <sup>2</sup> =0,979 AIC=134 RSE=2,18	R <sup>2</sup> =0,999 AIC=357 RSE=9,31	R <sup>2</sup> =0,991 AIC=359 RSE=7,66	R <sup>2</sup> =0,988 AIC=224 RSE=3,42	R <sup>2</sup> =0,991 AIC=249 RSE=3,79
<b>M6</b>	R <sup>2</sup> =0,989 AIC=118 RSE=1,61	R <sup>2</sup> =0,999 AIC=328 RSE=6,74	R <sup>2</sup> =0,993 AIC=347 RSE=6,68	R <sup>2</sup> =0,995 AIC=193 RSE=2,30	R <sup>2</sup> =0,996 AIC=211 RSE=2,43
<b>M7</b>	R <sup>2</sup> =0,977 AIC=137 RSE=2,27	R <sup>2</sup> =0,999 AIC=361 RSE=9,65	R <sup>2</sup> =0,991 AIC=356 RSE=7,45	R <sup>2</sup> =0,989 AIC=219 RSE=3,23	R <sup>2</sup> =0,990 AIC=252 RSE=3,94

Los criterios de calidad del ajuste estudiados se comportan de forma similar para cada modelo. El modelo logarítmico-logístico de cinco parámetros (M2) es seleccionado, por presentar mejores valores, en la provincia de Villa Clara, mientras que el modelo de dos fases de dosis-

respuesta de siete parámetros (M6) se presenta como mejor candidato para la Isla de la Juventud, La Habana, Ciego de Ávila y Matanzas. Sin embargo, no es posible calcular predicciones, con las técnicas utilizadas en esta investigación, para el pico de la provincia Ciego de Ávila, debido a



que el tiempo, al emplear el algoritmo para la búsqueda del punto de inflexión, sobrepasa la tolerancia empleada en la investigación sin haber obtenido resultado. En este caso, se selecciona el segundo modelo que proporciona los mejores índices de ajuste: M7.

La Tabla 3 presenta la prueba de bondad de ajuste para cada modelo seleccionado, sus

coeficientes e inferencias sobre estos y predicciones para picos de mayores contagios y cantidad total de estos, para cada uno de los territorios. Los coeficientes se nombran cuando se presentan sus valores ajustados y luego se muestran sus valores de significación al seguir el mismo orden en que se expusieron inicialmente.

**Tabla 3** - Coeficientes de modelos seleccionados y predicciones del día del pico de casos confirmados y total acumulado para los cinco territorios estudiados

Predicciones (02/05/2020)	Provincia				
	Isla de la Juventud	La Habana	Villa Clara	Ciego de Ávila	Matanzas
Modelo seleccionado	M6	M6	M2	M7	M6
Prueba de bondad del ajuste de Neill F (p-valor)	1,1083 (0,370 3)	1,8043 (0,846)	1,6415 (0,456)	1,732 (0,467)	0,3061 (0,1951)
Coeficientes del modelo	b1=-1,52 c=1,00 d1=47,66 a1=17,35 b2=-161,6 d2=8,72 a2=16,75	b1=-3,31 c=8,81 d1=1191 a1=42,78 b2=-35,18 d2=33217 a2=57,2	b=-30,2 c=4,67 d=190 a=38,8 h=0,11	b1=-2,33 b2=-4,89 c=2,61 d=83,7 a=14,9	b1=-9,22 c=1,11 d1=75,6 a1=38,4 b2=-2,19 d2=78,3 a2=19,2
Significación de los coeficientes	<0,001 0,58 <0,001 0,0016 <0,001 <0,001 <0,001	<0,001 <0,001 <0,001 <0,001 <0,001 <0,001 <0,001	<0,001 0,0081 <0,001 <0,001 0,0012	<0,001 <0,001 0,12 <0,001 <0,001	<0,001 0,50 <0,001 <0,001 <0,001 <0,001
Picos estimados (día)	15-22 de abril	25-30 de abril	11-17 de abril	4-6 de abril	26 de abril - 2 de mayo
Cantidad total de contagios estimados	42-57	1474-2514	190-1698	84-95	154-1238

Las pruebas de bondad de ajuste (mediante el Test de Neill), para medir la validez del ajuste,

aplicadas a cada territorio con cada uno de los 7 modelos dieron valores de significación mayores

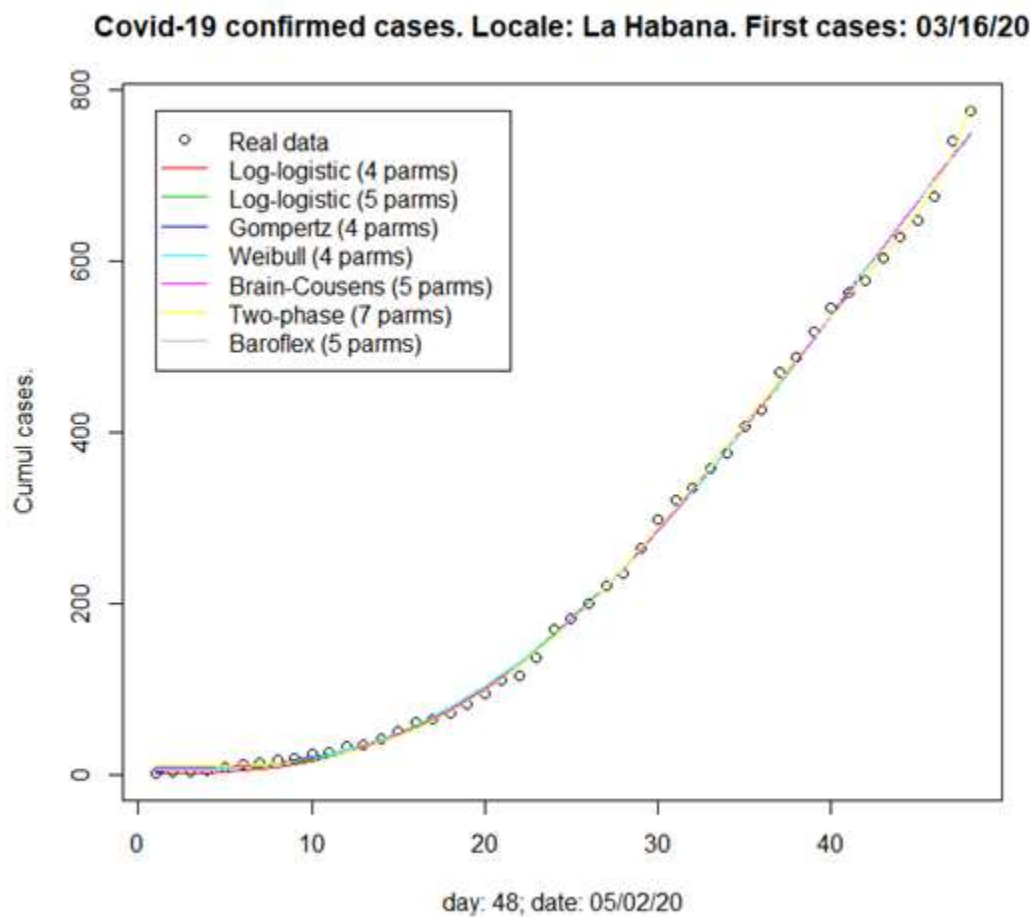




de 0,05 lo que significa que ninguno de los modelos se aparta significativamente de los datos empíricos, por lo que los 7 modelos aplicados son válidos para realizar pronósticos. En la Tabla 3 fueron mostrados solo los resultados correspondientes a los modelos seleccionados para cada territorio.

La Figura 1 muestra la representación geométrica de los casos confirmados acumulados y los 7 modelos contrastados para la provincia de La Habana. Los modelos se ajustan de forma que se acercan entre ellos. Esto pone de manifiesto, de forma gráfica, la aproximación entre los mismos y la adecuación hacia los datos.

**Figura 1-** Modelos de crecimiento poblacional ajustados a los casos confirmados acumulados en la provincia de La Habana



*Estudio de caso de Cuba*

En la Tabla 4 se presentan los resultados del procesamiento realizado para los casos confirmados a la COVID-19 en todo el país. Se

muestran los criterios de ajuste obtenidos para cada modelo.



**Tabla 4 - Criterios de ajuste para cada modelo procesado para los datos de toda Cuba**

Modelos para Cuba(02/05/2020)						
M1	M2	M3	M4	M5	M6	M7
R <sup>2</sup> =0,999	R <sup>2</sup> =0,999	R <sup>2</sup> =0,999	R <sup>2</sup> =0,988	R <sup>2</sup> =0,999	R <sup>2</sup> =0,999	R <sup>2</sup> =0,999
AIC=422	AIC=420	AIC=427	AIC=426	AIC=423	AIC=413	AIC=424
RSE=12,2	RSE=12	RSE=12,9	RSE=12,8	RSE=12,3	RSE=11	RSE=12,36

Los criterios de calidad del ajuste se comportan de forma similar para cada modelo. Se selecciona el modelo M6 que presenta los mejores valores. La prueba de Neill presenta niveles de significación mayores que 0,05 para cada uno de los 7 modelos aplicados a los datos de Cuba, lo

cual pone de manifiesto que los modelos son adecuados para realizar pronósticos. La Tabla 5 presenta los resultados de la prueba de Neill para el modelo M6, sus coeficientes ajustados y nivel de significación y predicciones para picos de mayores contagios y cantidad total de estos.

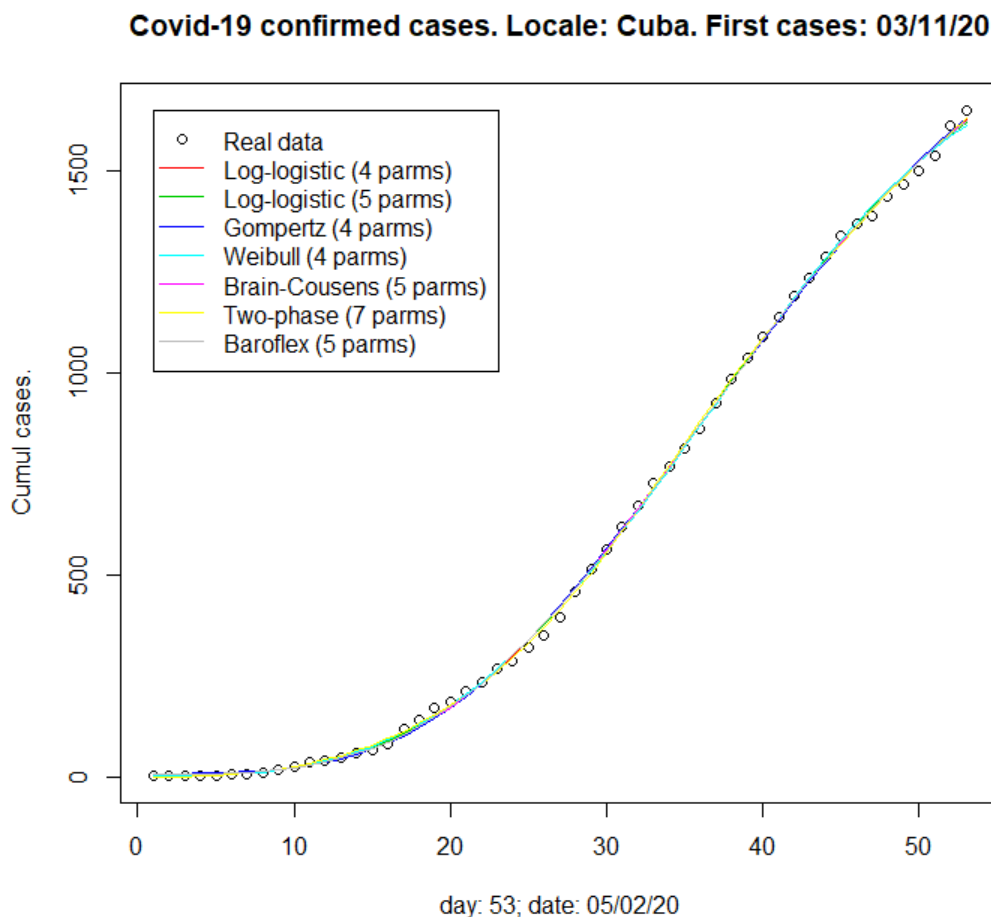
**Tabla 5 - Coeficientes del modelo seleccionado y predicciones del día del pico de casos confirmados y total acumulado de estos para toda Cuba**

Cuba (Modelo seleccionado: M6)				
Prueba de bondad del ajuste de Neill F (p-valor)	Coeficientes del modelo	Significación de los coeficientes	Estimados picos de mayor contagio(días)	Estimados cantidad de contagios
1,036 (0,456)	b1=-2,85 c=-0,42 d1=2547 a1=50,2 b2=-7,83 d2=265 a2=33,6	<0,001 0,92 <0,001 <0,001 <0,001 <0,001 <0,001	13-16 de abril	2140-2809

La Figura 2 muestra la representación geométrica de los casos confirmados acumulados en toda Cuba y los 7 modelos contrastados. Los modelos

se ajustan de forma adecuada a los datos y se aproximan entre ellos.



**Figura 2.** Modelos de crecimiento poblacional ajustados a los casos confirmados acumulados en Cuba

## DISCUSIÓN

En el presente estudio además de exponer predicciones para los días de mayor reporte de casos confirmados a la COVID-19 y cantidad total de estos para diferentes territorios de Cuba, se incluye a todo el país, se profundiza en la técnica utilizada para obtener dichos pronósticos. Investigaciones similares, realizadas, publicadas y revisadas por los autores han aplicado estas técnicas para obtener pronósticos sobre la epidemia provocada por la COVID-19 en otras

regiones del planeta, con diferentes niveles de profundización:

Dattoli<sup>(7)</sup> utiliza modelo logístico y Gompertz no aplica criterio para la selección de estos ni pruebas de significación. Jia,<sup>(4)</sup> Villalobos<sup>(10)</sup> y Tátrai-Várallyay<sup>(11)</sup> utilizan, indistintamente, modelos logísticos, de Gompertz y Bertalanffy y un único criterio para medir el ajuste sin aplicar pruebas de significación. Prueban la adecuación de los modelos con el acertado ajuste hacia datos de epidemias ya registrados.



Yang<sup>(5)</sup> contrasta modelos de crecimiento exponencial, modelos estadísticos de estimación y modelos SIR y SEIR; utiliza criterios de selección de modelos como: AIC, RSE e índice de robustez (RB); sin embargo, no presentan pruebas de significación sobre los modelos utilizados, por lo que no se conoce el nivel de confiabilidad de estos modelos, al igual que los casos anteriormente mencionados.

Batista<sup>(6)</sup> utiliza un criterio para evaluar el ajuste ( $R^2$ ), utiliza la prueba de significación F para probar la validez pero utiliza un único modelo: logístico.

En este estudio se utilizan 7 modelos, tres criterios para la selección del modelo de mejor ajuste y pruebas de significación, aspectos que resumen, de alguna manera, las investigaciones anteriormente mencionadas.

Los modelos utilizados en la presente investigación presentaron similares criterios sobre la calidad del ajuste y pasaron la prueba de bondad de ajuste. Esto permite su generalización para describir matemáticamente estas situaciones y pone de manifiesto la validez de cada uno para realizar pronósticos. No se puede decir basado en este estudio, que un modelo estadísticamente haya dado mejores resultados que el resto; sin embargo, se seleccionó uno de ellos para ser utilizado en cada territorio estudiado basado en los mejores valores de  $R^2$ , AIC y RSE.

El modelo M6, seleccionado para la Isla de la Juventud, Matanzas y toda Cuba junto con el modelo M7 seleccionado para Ciego de Ávila,

presentan el coeficiente  $c$  o  $c_1$  no significativo. Esto no invalida dichos modelos, el coeficiente  $c$  o  $c_1$  suele ser utilizado, en este contexto, como la cantidad inicial de casos al comenzar el estudio. Para el cálculo de resultados fueron tomados estos coeficientes, según los valores obtenidos al aplicar el ajuste; no obstante, debido a la función que cumplen en los modelos y que estos valores se acercan a cero, los resultados que se obtienen con estos coeficientes en cero serían muy próximos a los presentados en este estudio.

Los modelos utilizados para realizar predicciones describen correctamente los procesos de crecimientos poblacionales; aunque existen muchos factores que pueden influir en la propagación de una epidemia que en ocasiones la relación entre dos variables (dependiente e independiente) no puede expresarse.

La técnica utilizada está basada en las observaciones ya existentes y la obtención a partir de ellas de los parámetros de modelos ya preestablecidos. Mientras mayor sea el número de observación, mejor aproximación se obtendrá de los mismos; sin embargo, mayor tiempo habrá transcurrido de afectación de la epidemia.

Jia<sup>(4)</sup> ha realizado pronósticos satisfactorios, pasados aproximadamente 40 días luego de los primeros reportes y Villalobos<sup>(10)</sup> pasados 50 días. Los territorios analizados en esta investigación se encuentran pasados 40 días de haberse registrado los primeros casos a excepción de la Isla de la Juventud donde han transcurrido 29 días. Cuba se encuentra pasados 53 días.



## CONCLUSIONES

Los modelos de crecimiento poblacional para la obtención de pronósticos de días de mayor afectación y cantidad total de contagios, por la COVID-19 en territorios de Cuba y en todo el país, han sido probados estadísticamente y dan resultados significativos. Los mismos presentan validez y criterios de ajuste similares por lo que se puede concluir que todos los modelos utilizados en el estudio presentan adecuación para realizar

pronósticos.

En la presente investigación se obtuvieron predicciones al utilizar los modelos M2, M6 y M7, luego de probar la validez del ajuste mediante prueba de bondad de ajuste y de significación sobre sus coeficientes. De esta forma se evidencia la confiabilidad en las predicciones presentadas.

## RREFERENCIAS BIBLIOGRÁFICAS

1. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis.* 2020;12(3):165.
2. Cagigal MAG, Becario FPU. Modelado y análisis de la evolución de una epidemia vírica mediante filtros de Kalman: el caso de la COVID-19 en España. En: Depósito de investigación de la Universidad de Sevilla [Internet]. España: Universidad de Sevilla; 2020 [Citado 13/05/2020]. Disponible en: [https://idus.us.es/bitstream/handle/11441/94508/Analisis%20de%20la%20evolucion%20del%20COVID-19%20con%20Filtros%20de%20Kalman%20v1%20%20801-04-2020%29.pdf?sequence=7&isAllowed=yLo\\_guarde\\_pdf](https://idus.us.es/bitstream/handle/11441/94508/Analisis%20de%20la%20evolucion%20del%20COVID-19%20con%20Filtros%20de%20Kalman%20v1%20%20801-04-2020%29.pdf?sequence=7&isAllowed=yLo_guarde_pdf)
3. Simón Mínguez F. Procesos de difusión Logístico y Gompertz. Métodos numéricos clásicos en la estimación paramétrica [Tesis de maestría]. España: Universidad de Granada; 2016 [Citado 13/05/2020]. Disponible en: [https://masteres.ugr.es/moea/pages/curso201516/tfm1516/simon\\_minguez\\_tfm/](https://masteres.ugr.es/moea/pages/curso201516/tfm1516/simon_minguez_tfm/)
4. Jia L, Li K, Jiang Y, Guo X. Prediction and analysis of Coronavirus Disease 2019 [Internet]. Nueva York: ArXiv; 2020 [Citado 13/05/2020]. Disponible en: <https://arxiv.org/abs/2003.05447>
5. Yang W, Zhang D, Peng L, Zhuge C, Hong L. Rational evaluation of various epidemic models based on the COVID-19 data of China [Internet]. Nueva York: ArXiv; 2020 [Citado 13/05/2020]. Disponible en: <http://arxiv.org/abs/2003.05666>
6. Batista M. Estimation of the final size of the COVID-19 epidemic [Internet]. Estados Unidos: MedRxiv; 2020 [Citado 13/05/2020]. Disponible en: <https://www.medrxiv.org/content/10.1101/2020.02.16.20023606v5>
7. Dattoli G, Di Palma E, Licciardi S, Sabia E. A Note on the Evolution of Covid-19 in Italy [Internet]. Nueva York: ArXiv; 2020 [Citado 13/05/2020]. Disponible en: <http://arxiv.org/abs/2003.08684>
8. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next?. *The Lancet* [Internet]. 2020 Abr [Citado 13/05/2020];395:1225-8. Disponible en: <https://www.thelancet.com/article/S0140-6736%2820%2930627-9/abstract>
9. Qeadan F, Honda T, Gren LH, Dailey Provost J, Benson LS, Van Derslice JA, et al. Naive Forecast for COVID-19 in Utah Based on the South Korea and Italy Models-the Fluctuation between Two Extremes. *Int J Environ Res Public Health* [Internet]. 2020 [Citado 13/05/2020];17(8):2750. Disponible en: <https://www.ncbi.nlm.nih.gov/pubmed/32316165>



10. Villalobos-Arias M. Estimation of population infected by Covid-19 using regression Generalized logistics and optimization heuristics [Internet]. Nueva York: ArXiv; 2020 [Citado 13/05/2020]. Disponible en: <https://arxiv.org/abs/2004.01207>
11. Tátrai D, Várallyay Z. COVID-19 epidemic outcome predictions based on logistic fitting and estimation of its reliability [Internet]. Nueva York: ArXiv; 2020 [Citado 13/05/2020]. Disponible en: <https://arxiv.org/pdf/2003.14160>
12. Seber GAF. Nonlinear Regression Models. En su: The Linear Model and Hypothesis: A General Unifying Theory. Cham [Internet]. Nueva York: Springer International Publishing; 2015 p.117-28 [Citado 13/05/2020]. Disponible en: <https://www.springer.com>
13. Moamer S, Baghestani AR, Pourhoseingholi MA, Maboudi AAK, Shahsavari S, Zali MR, et al. Application of the parametric regression model with the four-parameter log-logistic distribution for determining of the effecting factors on the survival rate of colorectal cancer patients in the presence of competing risks. Iran Red Crescent Med J [Internet]. 2017 [Citado 13/05/2020];19(6):[aprox. 2 p.]. Disponible en: <https://sites.kowsarpub.com/ircmj/articles/13185.html>
14. Finney DJ. Bioassay and the practice of statistical inference. Int Stat Rev Int Stat [Internet]. 1979 [Citado 13/05/2020];47(1):1-12. Disponible en: <https://www.sciencedirect.com/science/article/pii/0169260796017257/>
15. Kyurkchiev N, Iliev A. Extension of Gompertz-type Equation in Modern Science: 240 Anniversary of the birth of B. Gompertz. [Internet]. Deutschland: LAP LAMBERT Academic Publishing; 2018 [Citado 13/05/2020]. [https://www.researchgate.net/publication/327137112\\_Extension\\_of\\_Gompertz-](https://www.researchgate.net/publication/327137112_Extension_of_Gompertz-type_Equation_in_Modern_Science_240_Anniversar_y_of_the_birth_of_B_Gompertz)
16. Wais P. A review of Weibull functions in wind sector. Renew Sustain Energy Rev. 2017;70:1099-1107.
17. Brain P, Cousens R. An equation to describe dose responses where there is stimulation of growth at low doses. Weed Res. 1989;29(2):93-6.
18. Groot JCJ, Cone JW, Williams BA, Debersaques FMA, Lantinga EA. Multiphasic analysis of gas production kinetics for in vitro fermentation of ruminant feeds. Anim Feed Sci Technol. 1996;64(1):77-89.
19. Ricketts JH, Head GA. A five-parameter logistic equation for investigating asymmetry of curvature in baroreflex studies. Am J Physiol-Regul Integr Comp Physiol. 1999;277(2):R441-R454.
20. R: The R Project for Statistical Computing [Internet]. Austria: R Foundation for Statistical Computing; 2018. [Citado 13/05/2020]. Disponible en: <https://www.r-project.org/>
21. Colin Cameron A, Windmeijer FAG. An R-squared measure of goodness of fit for some common nonlinear regression models. J Econom. 1997;77(2):329-42.
22. Osuji GA, Okoro CN, Obubu M, Obiora Llouno HO. Effect of akaike information criterion on model selection in analyzing auto-crash variables. Int J Sci Basic Appl Res. 2016;26(1):98-109.
23. Draper NR, Smith H. Applied regression analysis [Internet]. 3ra ed. United States: John Wiley & Sons; 1998 [Citado 13/05/2020]. Disponible en: <https://www.wiley.com/en-us/Applied+Regression+Analysis%2C+3rd+Edition-p-9780471170822>
24. Neill JW. Testing for lack of fit in nonlinear regression. Ann Stat. 1988;16(2):733-40.



**Conflicto de intereses**

Los autores declaran no tener ningún conflicto de intereses.

**Contribución de autoría**

**JFMM:** análisis estadístico, escritura del artículo y cálculo de resultados.

**MECC:** análisis de documentos, escritura del artículo y revisión general.

**MCI:** análisis de documentos, escritura del artículo y revisión general.

Todos los autores participamos en la discusión de los resultados y hemos leído, revisado y aprobado el texto final del artículo.

