

# Aplicación de modelos de crecimiento sigmoideal en el pronóstico en tiempo real del brote de la Covid-19 en Villa Clara

## Application of sigmoidal growth models in real-time forecast of Covid-19 outbreak in Villa Clara

Osvaldo Norman Montenegro<sup>1\*</sup>, David Rodríguez Fernández<sup>2</sup>, Carlos de la Caridad Rodríguez Fadrugas<sup>2</sup>, Carlos Sebrango Rodríguez<sup>3</sup>, Yanet Rodríguez Sarabia<sup>4</sup>, Roxana Pérez García<sup>5</sup>, Luis B. Lecha Estela<sup>6</sup>, María del Carmen Velasco Gómez<sup>7</sup>, Bárbara Rodríguez<sup>8</sup>

**Resumen** En el presente artículo se estudia la dinámica de transmisión del SARS CoV-2 en Villa Clara utilizando diferentes modelos matemáticos, entre ellos los modelos fenomenológicos de crecimiento sigmoideal y de dinámica de población SIR. A partir del ajuste de los modelos a las series de datos de infectados acumulados, diariamente se evaluaron y estimaron diferentes parámetros asociados a la marcha de la Covid-19 en la provincia, información considerada por las autoridades en la toma de decisiones. Se presenta además, en retrospectiva, la evaluación de los pronósticos realizados en momentos representativos de las tres etapas identificadas, avalándose la intervención oportuna de las autoridades sanitarias para controlar eventos de transmisión local suscitados en el territorio.

**Abstract** The present paper studies the transmission dynamics of SARS CoV-2 in Villa Clara using different mathematical models, including phenomenological models of sigmoidal growth and SIR population dynamics. Based on the adjustment of the models to the series of accumulated infected data, different parameters associated with the progress of Covid-19 in the province were evaluated and estimated daily, information considered by the authorities in the decision making process. The evaluation of the forecasts made at representative moments of the three identified stages is also presented in retrospect, supporting the timely intervention of the health authorities to control the local transmission event that occurred in the territory.

### Palabras Clave

Covid-19, modelos de crecimiento sigmoideal, SIR, pronósticos en tiempo real

### Keywords

Covid-19, sigmoidal growth models, SIR, real-time forecast

<sup>1</sup> Centro de Bioactivos Químicos, Villa Clara, Cuba, onorman@uclv.edu.cu

<sup>2</sup> Departamento de Física, Universidad Central Marta Abreu de Las Villas, Villa Clara, Cuba, davidrf@uclv.cu, fadrugas@uclv.edu.cu

<sup>3</sup> Centro de Estudios de Energía y Procesos Industriales CEEPI, Sancti Spiritus, Cuba, sebrango@uniss.edu.cu

<sup>4</sup> Departamento de Ciencias de la Computación, Universidad Central Marta Abreu de Las Villas, Villa Clara, Cuba, yrsarabia@uclv.edu.cu

<sup>5</sup> Departamento de Matemática, Universidad Central Marta Abreu de Las Villas, Villa Clara, Cuba

<sup>6</sup> Centro Meteorológico de la Isla de la Juventud, INSMET, Isla de la Juventud, Cuba, luis.lecha@gmail.com

<sup>7</sup> Delegación del CITMA en Villa Clara, Villa Clara, Cuba, mcv@citmavcl.gob.cu

<sup>8</sup> Centro Provincial de Higiene, Epidemiología y Microbiología, Villa Clara, Cuba, barbaritara@infomed.sld.cu

\*Autor para Correspondencia, Corresponding Author

## Introducción

Las enfermedades infecciosas como malaria, tuberculosis, VIH/SIDA, síndrome respiratorio agudo grave (SRAG) han provocado en años recientes efectos económicos y de salud desfavorables. Por ello, el uso de métodos cuantitativos, basados en modelos matemáticos para estudiar la dinámica de transmisión y control de estas enfermedades ha ganado impor-

tancia de forma notoria entre los científicos y profesionales de la salud con el objetivo de proponer programas efectivos de control e interpretar patrones epidemiológicos [14]. El nuevo coronavirus tipo SARS (Severe Acute Respiratory Syndrome), nombrado como SARS CoV-2 [9], causante de la Pandemia de la Covid-19 [17], ha generado crisis en materia sanitaria, social y económica sin importar el nivel desarrollo de los países;

y demanda para su combate efectivo una mayor vinculación y participación de los gobiernos y comunidades científicas.

En [6] se reflexiona sobre esta experiencia en Cuba, mostrando el sistema de trabajo utilizado y las principales acciones realizadas, destacando el relevante papel que han desempeñado la ciencia y la tecnología nacionales. Los primeros casos de la Covid-19 se detectan en Cuba el 11 de marzo y progresivamente la situación se va complejizando en el país. A partir del día 21 de marzo, la dirección del gobierno decide activar los Consejos de Defensa Provinciales (CDP), y el 24 de marzo la Presidenta del CDP de Villa Clara solicita a la Jefa del Subgrupo de Ciencias conformar un grupo de investigadores que trabajaran en un pronóstico o estimación del comportamiento de la Covid-19 en el territorio. Este grupo queda conformado por: físicos, matemáticos, meteorólogos, geógrafos, graduados de ciencias de la computación y epidemiólogos. Su principal misión era asistir al consejo de defensa provincial brindando pronósticos acertados sobre los principales indicadores, y que les permitiera a las autoridades sanitarias mantener un control adecuado sobre el desarrollo de la epidemia en la provincia.

El enfrentamiento a la Covid-19 en Villa Clara, como en todo el país, ha exigido por las autoridades de salud y el CDP un seguimiento de variables tales como: cantidad de personas hospitalizadas y en centros de aislamiento, el tamaño total del brote (total de casos confirmados), y el pico de incidencia (día donde se alcanza el mayor número de contagiados). El objetivo del presente trabajo es mostrar el uso de diferentes modelos matemáticos reportados en la literatura, entre estos los modelos fenomenológicos de crecimiento sigmoïdal y el conocido modelo SIR, a los efectos de evaluar y estimar el comportamiento de los diferentes parámetros asociados a la marcha de la Covid-19 en Villa Clara. Además, se presenta la validación de los pronósticos realizados desde de este grupo, conocido en el territorio como Grupo Científico Asesor del CDP de Villa Clara (GCA-CDP), distinguiendo entre el 11 de marzo y el 18 de mayo tres etapas fundamentales del brote que arrojó un total de 216 casos confirmados.

## 1. Materiales y Métodos

Se realizó un estudio predictivo utilizando varios modelos fenomenológicos de crecimiento sigmoïdal y el modelo SIR. Para cada uno de los modelos la variable dependiente resultó el número de casos confirmados acumulados a la Covid-19 y el número de contagiados activos en el caso del modelo SIR, mientras que la variable independiente fue determinada por el día de evolución de la enfermedad a partir del primer caso confirmado en la provincia, el 11 de marzo de 2020. El período de tiempo analizado en el estudio, correspondió al intervalo entre el 11 de marzo y el 18 de mayo de 2020, fecha en la cual fue considerada el fin de la epidemia en Villa Clara por el GCA\_CDP, según los criterios epidemiológicos considerados como la no ocurrencia de casos durante 2 períodos de incubación del virus, es decir durante 28 días.

Las series de datos analizadas fueron obtenidas a partir de

los reportes oficiales emitidos por parte de las autoridades de salud pública y epidemiología de la provincia de Villa Clara. Para el tratamiento de los datos y el procesamiento estadístico se utilizaron los softwares Wolfram Mathematica versión 11.0 del 2016 y el R versión 4.0 del 2020.

Debido a la dinámica de trabajo del grupo, las series de datos de contagiados acumulados y activos fueron analizadas diariamente, se generaron pronósticos de las variables dependientes y estimados de los parámetros característicos de cada modelo.

### 1.1 Modelos fenomenológicos de crecimiento sigmoïdal

En 1959, en su artículo "A flexible growth function for empirical use?", Richards F. J. [20] propone una generalización al modelo de von Bertalanffy para describir el incremento en peso,  $W(t)$ , de una determinada especie animal. Se asume que el ritmo de crecimiento en peso de la especie es proporcional a la diferencia entre el ritmo anabólico y el ritmo catabólico. Según plantea Pütter [19] el ritmo anabólico es proporcional a la  $m$ -ésima potencia del peso,  $W(t)^m$ , y el ritmo catabólico es una función lineal del peso. Lo anterior, conduce a escribir la siguiente ecuación diferencial.

$$\frac{dW(t)}{dt} = -\kappa W(t)(1 - (\eta/\kappa)W(t)^{m-1}), \quad (1)$$

donde  $\eta$  y  $\kappa$  son las constantes de anabolismo y catabolismo respectivamente. La ecuación diferencial (1) es el conocido modelo de Richards. El índice exponencial  $m$ , es un parámetro empírico de ajuste, toma valor 2/3 para el modelo de von Bertalanffy, en caso de organismos que obedecen la regla superficial de metabolismo. Como se mostrará posteriormente, el modelo de Richards no solo generaliza al modelo de von Bertalanffy, sino a un grupo de modelos de crecimiento sigmoïdal, entre los que se destacan: el modelo exponencial negativo, el modelo logístico y el modelo de Gompertz entre otros.

El modelo de Richards no ha sido utilizado exclusivamente en el crecimiento en peso de especies animales. Durante el estudio de la dinámica de brotes de enfermedades infecciosas, dicho modelo se ha empleado, para describir y pronosticar el comportamiento de variables dinámicas acumulativas, tales como la población de contagiados acumulados, la población de fallecidos y la población de recuperados [8, 18, 26]. En el campo de la epidemiología la ecuación diferencial de Richards suele escribirse, similar a la ecuación (1), en la forma:

$$\frac{dG(t)}{dt} = rG(t)((1 - (G(t)/K)^a), \quad (2)$$

donde  $G(t)$  representa la variable dinámica acumulativa,  $r$  es el ritmo de contacto efectivo,  $K$  es el tamaño total del brote epidémico o capacidad de carga del sistema y  $a$ , relacionada directamente con  $m$  mediante  $a = m - 1$ , es un parámetro empírico de ajuste que agrega complejidad al modelo y garantiza flexibilidad a la solución  $G(t)$ .

Teniendo en cuenta la condición inicial  $G(0) = G_0$ , la ecuación (2) puede ser fácilmente integrada, obteniéndose que:

$$G(t) = K [1 + S \exp(-\rho t)]^{-1/a}, \quad (3)$$

donde  $S = (K/G_0)^a - 1$  y  $\rho = |ra|$ . Para valores de  $a > -1$ ,  $G(t)$  representa una función de crecimiento sigmoide [22], cuyo punto de inflexión tiene coordenadas  $(t_i = -\ln(S/a)/\rho, G(t_i) = K [1 + a]^{-1/a})$ . Reparametrizando la ecuación (3) en términos del tiempo de inflexión  $t_i$ , se obtiene:

$$G(t) = K [1 + a \exp(-\rho(t - t_i))]^{-1/a}, \quad (4)$$

para valores de  $-1 < a < 0$ , la ecuación (4) intercepta al eje del tiempo en  $t_0 = \ln(-S)/\rho$ , se dice que dicho modelo tiene un punto de inicio. Al contrario, para valores de  $a > 0$ ,  $G(t)$  presenta una asíntota horizontal inferior en  $G = 0$ . En la figura 1 se observa una representación gráfica de la función  $G(t)$  para diferentes valores del parámetro  $a$ , manteniéndose constantes los demás parámetros. Cada uno de los valores escogidos para  $a$  se asocian a un modelo específico, como caso particular del modelo de Richards, posteriormente se describirán en detalles.

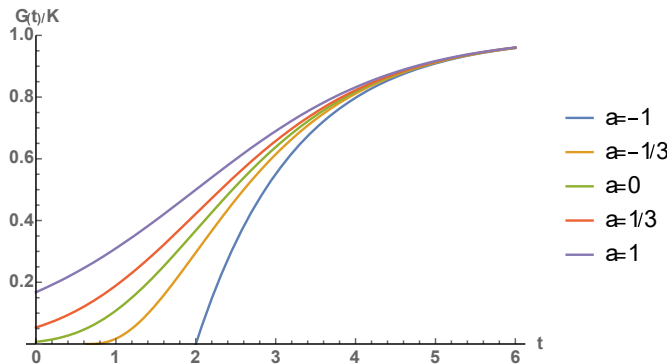


Figura 1. Representación gráfica de  $G(t)$  para valores de  $a = \{-1, -1/3, 0, 1/3, 1\}$

La ecuación (4) es una de las formas más utilizadas, en el campo de la epidemiología del modelo de Richards de cuatro parámetros  $\{K, a, \rho, t_i\}$ . Si la variable dinámica  $G(t)$  del modelo se utiliza para describir la población de individuos contagiados acumulados, entonces mediante un ajuste a la serie de datos reales puede obtenerse información vital para las autoridades de salud, dígase: el número máximo de individuos contagiados, el ritmo de contacto efectivo (asociado con el ritmo de transmisión de la enfermedad) y el tiempo al cual ocurre el máximo de incidencia.

### 1.1.1 Modelo exponencial negativo

Sea  $a = -1$ , el modelo de Richards degenera en el modelo exponencial negativo, representado en la figura 1 por la curva azul. Luego, el modelo exponencial negativo se expresa mediante la ecuación:

$$G(t) = K [1 - \exp(-r(t - t_i))]. \quad (5)$$

Es válido indicar que este modelo no representa una curva sigmoide. Es una curva de crecimiento acotado, con punto de inicio en la coordenada  $t_i = t_0 = \ln(-S)/\rho$  y asíntota horizontal superior en  $G = K$ . Suele utilizarse en el ajuste de series de datos después de superado el punto de inflexión, en la historia final del brote epidémico.

### 1.1.2 Modelo de von Bertalanffy

Como ya se había indicado, el modelo de von Bertalanffy se obtiene haciendo  $a = -1/3$  en el modelo de Richards. En la figura 1, la curva naranja representa dicho modelo. En este caso, la ordenada del punto de inflexión es  $G(t_i) = 2/3^3 K \approx 0,296K$ , es decir: el punto de inflexión se obtiene en el 29,6% de la capacidad de carga del sistema. La ecuación del modelo de von Bertalanffy se expresa como:

$$G(t) = K [1 - (1/3) \exp(-r(t - t_i)/3)]^3. \quad (6)$$

### 1.1.3 Modelo Logístico o de Verhulst

El modelo Logístico, propuesto en 1838 por Verhulst P.F. [24] como crítica a la ley de crecimiento exponencial propuesta por Thomas Malthus, puede ser obtenido como caso particular del modelo de Richards cuando  $a = 1$ . Ha sido ampliamente utilizado para describir el crecimiento de poblaciones [13, 16, 27], y en especial para estudiar la dinámica de crecimiento de brotes epidémicos [15, 28]. Se representa por la curva superior, de color violeta, en la figura 1 y viene dado por la ecuación:

$$G(t) = \frac{K}{1 + \exp(-r(t - t_i))}. \quad (7)$$

Para valores pequeños del tiempo la función logística puede ser aproximada a la ley de crecimiento exponencial  $G(t) = G_0 \exp(rt)$ , ampliamente utilizada para estudiar los momentos iniciales de un brote epidémico [4]. Además, en las primeras etapas, la ley de crecimiento exponencial permite estimar del número básico de reproducción del brote,  $R_0$  [25], mediante la ecuación:

$$R_0 = 1 + rD, \quad (8)$$

donde  $D$  es el tiempo medio infeccioso.

### 1.1.4 Modelo de Gompertz

La obtención del modelo de Gompertz como caso particular del referido modelo de Richards, no es tan simple como los casos anteriores. Se plantea [20] que el modelo de Gompertz resulta del paso al límite cuando  $a \rightarrow 0$ . Reescribiendo la ecuación (2), en la forma:

$$\frac{dG(t)}{dt} = \rho G(t) ((1 + (G(t)/K)^a) / a), \quad (9)$$

y teniendo en cuenta que el  $\lim_{a \rightarrow 0} (1 - x^a) / a = -\ln(x)$ , se obtiene que:

$$\frac{dG(t)}{dt} = \rho G(t) \ln(K/G(t)). \quad (10)$$

La ecuación (10) es conocida como ecuación diferencial de Gompertz y mediante integración directa se obtiene la solución en la forma:

$$G(t) = K \exp(-\exp(-\rho(t-t_i))). \quad (11)$$

En este caso, la ordenada del punto de inflexión es  $G(t_i) = K/e$ . El punto de inflexión se alcanza aproximadamente al 36,8 % de la capacidad de carga. El modelo de Gompertz presenta un crecimiento inicial más precipitado en comparación con la segunda etapa del proceso, luego de superado el punto de inflexión. Este modelo fue propuesto por Benjamín Gompertz [7], para ajustar el incremento en el ritmo de muertes con la edad de los individuos. Los resultados obtenidos fueron rápidamente aplicados por empresas de seguros, para estimar el riesgo de muertes. Posteriormente, en su forma acumulativa conocida como Modelo de Gompertz-Makeham [12], ha sido ampliamente aplicado al crecimiento en peso y tamaño de diferentes especies animales como: aves, reptiles y mamíferos, así como al estudio y modelación del crecimiento de tumores y la supervivencia de pacientes con cáncer [22]. Además, varios son los trabajos [21, 23], que utilizan el modelo de Gompertz en el ajuste de variables acumulativas asociadas al campo de la epidemiología, como la transmisión de enfermedades infecciosas.

### 1.2 Modelo Epidemiológico SIR

De manera similar, la aplicación de la teoría de dinámica de poblaciones al estudio de brotes epidémicos ha permitido el desarrollo de modelos deterministas con enfoque compartimental. Estos modelos pueden alcanzar un alto nivel de complejidad, asociado fundamentalmente al número de variables dinámicas y parámetros que contengan [2, 5]. Sin embargo modelos clásicos y más simples como el SIR, formulado por A.G. MacKendrick y W.O. Kermack en 1927 [11], resultan sumamente valiosos. El modelo SIR separa una determinada población en tres compartimentos fundamentales, ver figura 2: Susceptibles,  $S(t)$ : individuos de la población susceptibles a ser infectados, Infectados,  $I(t)$ : Individuos infectados y a su vez infecciosos, y Removidos,  $R(t)$ : individuos que han sido infectados y posteriormente se han recuperado o fallecido a causa de la enfermedad. La población total es cerrada y tiene un tamaño efectivo  $N = S(t) + I(t) + R(t)$ , por lo general es menor al tamaño de la población real bajo estudio.

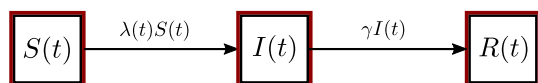


Figura 2. Diagrama de transferencia del modelo epidemiológico SIR

En la función de transferencia entre las subpoblaciones de individuos susceptibles e infectados, el término  $\lambda(t)$  se conoce como fuerza de infección. Si se asume que los individuos se relacionan de manera caótica y desordenada, como moléculas que realizan un movimiento browniano, entonces

la fuerza de infección es una función lineal de los individuos infectados  $\lambda(t) = \beta I(t)$ , donde  $\beta$  es el ritmo de contacto efectivo. En el caso anterior, también suele decirse que se cumple la ley de acción de masas. Luego, si el período de permanencia de individuos en el compartimento de infectados sigue una distribución exponencial, entonces  $\gamma$  es una constante denominada ritmo de removimiento y su inverso es el tiempo medio infeccioso,  $D$ . Bajo estas suposiciones, el modelo SIR se representa mediante un sistema no lineal de ecuaciones diferenciales ordinarias de la forma:

$$\frac{dS(t)}{dt} = -r \frac{I(t)S(t)}{N}, \quad (12)$$

$$\frac{dI(t)}{dt} = r \frac{I(t)S(t)}{N} - \gamma I(t), \quad (13)$$

$$\frac{dR(t)}{dt} = \gamma I(t), \quad (14)$$

donde  $r$ , como mismo se definió para los modelos de crecimiento sigmoial, es el ritmo de contacto efectivo per cápita, y viene dado por  $\beta N$ . Un parámetro que puede obtenerse de manera indirecta a partir de  $r$  y  $\gamma$ , es el número de reproducción efectivo  $R_0 = r/\gamma$ , definido como el número de segundos contagios que introduce un individuo infectado en una población de individuos susceptibles. Este parámetro resulta de gran interés para las autoridades epidemiológicas. Si  $R_0 \leq 1$ , entonces el número de individuos infectados decaerá rápidamente a cero. De otra manera, si  $R_0 > 1$ , entonces el número de individuos infectados comenzará a aumentar alcanzando un pico y luego irá tendiendo a cero cuando  $t \rightarrow \infty$ . Este escenario de aumento en el número de contagiados se denomina brote epidémico. A partir del  $R_0$ , se puede determinar el número de reproducción efectivo  $R_e(t) = R_0 S(t)/N$ . Este nuevo parámetro resulta sumamente útil a la hora de medir el impacto de las medidas gubernamentales de control sobre el brote de la enfermedad infecciosa.

### 1.3 Ajuste, selección y evaluación de los pronósticos

El comportamiento de los modelos fue analizado en dos etapas diferentes. Una primera etapa, de calibración, en la cual se verificó la bondad del ajuste a los datos mediante criterios estadísticos, y se aplicó un criterio de selección de modelos basado en criterios de información. Así como una segunda etapa, de evaluación del desempeño en el pronóstico, a partir de métricas estadísticas de rendimiento.

#### 1.3.1 Etapa de calibración

Para valorar la bondad de ajuste de los modelos y compararlos entre sí, se tuvieron en cuenta los siguientes criterios:

1. El coeficiente de determinación  $R^2$ , el cual es adecuado para modelos no lineales y evalúa la efectividad que presenta la variable independiente para medir la variable dependiente.



- Inspección visual de la trayectoria de cada modelo mediante sus gráficos correspondientes, chequeando la no presencia de observaciones anómalas.
- Criterio de información de Akaike ( $AIC$ ): Escoge como función de mejor ajuste aquella que haga mínima la pérdida de información a partir de hacer mínimo el índice  $AIC$ . Cuando la razón entre el total de datos  $n$ , y el número de parámetros del modelo  $p$  no excede el valor de 30, Hurvich et al. [10] introducen una corrección de segundo orden al criterio de Akaike, expresada por:

$$AIC_c = AIC + \frac{2p(p+1)}{n-(p+1)}. \quad (15)$$

Los valores individuales de  $AIC$  y  $AIC_c$  no brindan, por si solos, mucha información debido a que contienen constantes arbitrarias y dependen del tamaño de la muestra [1](se han obtenido valores de  $AIC$  desde  $-600$  hasta  $340000$ ). La diferencia  $\Delta AIC_i = AIC_i - AIC_{\min}$ , donde  $AIC_{\min}$  es el mínimo del vector formado por los  $AIC_i$  obtenidos para cada modelo, resulta un término más valioso a la hora de comparar los modelos. Luego,  $\Delta AIC_i$  representa la pérdida de información experimentada por utilizar como modelo de ajuste aquel con índice de Akaike  $AIC_i$ , en vez del modelo más probable. Otro término sumamente útil a la hora de comparar los modelos son las pesadas de Akaike,  $w_i$ .

$$w_i = \frac{\exp(-\Delta AIC_i/2)}{\sum_{r=1}^R \exp(-\Delta AIC_r/2)}, \quad (16)$$

donde  $R$  es el número total de modelos candidatos. El término  $w_i$  se interpreta como la probabilidad de que el  $i$ -ésimo modelo sea más probable entre los posibles candidatos.

### 1.3.2 Etapa de pronóstico

Para cuantificar los errores asociados con los pronósticos se tuvieron en cuenta las métricas de rendimiento, presentadas por Chowell en [3], estas son:

- Raíz cuadrada del error cuadrático medio, (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \varepsilon(t_i)^2}, \quad (17)$$

donde  $\varepsilon(t_i) = f(t_i, \hat{\theta}) - y_i$  es la diferencia entre el modelo de mejor ajuste y la serie de datos temporales como función del tiempo.

- Medida del error absoluto, (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |\varepsilon(t_i)|. \quad (18)$$

- Media del error porcentual absoluto, (MAPE):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\varepsilon(t_i)|}{y_i}. \quad (19)$$

Cada una de estas métricas son una medida cuantitativa de la dispersión del comportamiento real de los datos respecto al pronóstico realizado. De esta manera, se considera que el modelo que logre minimizar el valor de las métricas presenta un mayor rendimiento ante el pronóstico.

## 2. Resultados y Discusión

A continuación, se reflejan en tres momentos diferentes los resultados obtenidos en el análisis de la evolución de la Covid-19 en la provincia de Villa Clara. Los momentos seleccionados fueron los siguientes: un primer momento constituido por el informe presentado al CDP el lunes 30 de marzo, donde se analiza la fase temprana del brote en la provincia la cual recoge los acontecimientos ocurridos entre el 26 y 30 de marzo del 2020, a raíz de los eventos que se sucedieron en la región nordeste de la provincia, que incluyeron los casos de Camajuaní y Caibarién. Un segundo momento constituido por el informe realizado el lunes 20 de abril, relacionado con los acontecimientos ocurridos en el hogar de ancianos en Santa Clara, entre el 15 y el 18 de abril de 2020 los cuales prácticamente duplicaron la cantidad de casos confirmados existentes hasta ese momento en la provincia; y por último el final de la contienda, es decir el informe del lunes 24 de mayo, informe a partir del cual no se reportaron más casos en Villa Clara.

### 2.1 Etapa inicial, del 11 de marzo al 29 de marzo

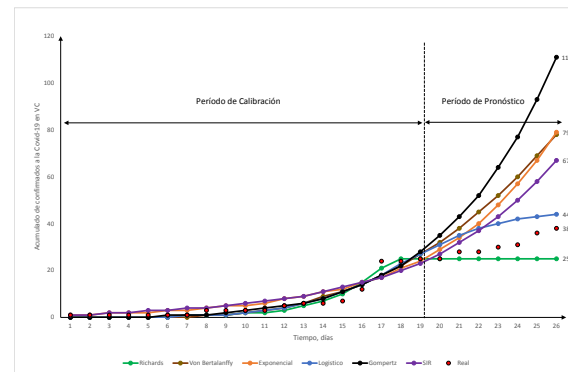


Figura 3. Ajuste de los modelos a la serie de datos de contagiados acumulados hasta el día 29 de marzo.

En esta etapa temprana del brote epidémico, se ajustaron los modelos propuestos a la serie de datos temporales, de dimensión 19. Además, se incorpora al análisis la ley de crecimiento exponencial. La tabla 1 presenta los resultados obtenidos mediante el ajuste, dígame: el coeficiente de determinación como medida de la bondad del ajuste y los parámetros estimados con su error estándar correspondiente. Todos los modelos presentan un ajuste aceptable a los datos con valores de  $R^2$  superiores a 0,9. El modelo de Richards resultó el

**Tabla 1.** Parámetros estimados y coeficiente de regresión obtenidos a partir del ajuste de los modelos a los datos de contagiados acumulados hasta los días 29 de marzo, 19 de abril y 23 de marzo

Fecha	Modelos	$K$	$t_i$	$a$	$r$	$R^2$
29 de marzo	Exponencial	—	—	—	$0,168 \pm 0,004$	0,925
	Richards	$24 \pm 1$	$17 \pm 9 \times 10^5$	$116 \pm 1,2 \times 10^9$	$0,37 \pm 7,7 \times 10^6$	0,978
	von Bertalanffy	$1,6 \times 10^4 \pm 2,8 \times 10^5$	$133 \pm 842$	$-1/3$	$0,008 \pm 0,06$	0,957
	Logístico	$46 \pm 22$	$18 \pm 2$	1	$0,4 \pm 0,1$	0,961
	Gompertz	$1,6 \times 10^4 \pm 2 \times 10^5$	$72 \pm 150$	$\rightarrow 0$	$0,03 \pm 0,06$	0,965
	SIR	$3,8 \times 10^6 \pm 8,5 \times 10^{10}$	—	—	$0,21 \pm 0,01$	0,908
19 de abril	Richards	$168 \pm 5$	$38 \pm 1$	$51 \pm 150$	$0,1 \pm 0,7$	0,994
	von Bertalanffy	$7 \times 10^4 \pm 4,7 \times 10^5$	$287 \pm 686$	$-1/3$	$0,004 \pm 0,009$	0,990
	Logístico	$797 \pm 488$	$50 \pm 7$	1	$0,12 \pm 0,01$	0,993
	Gompertz	$1,4 \times 10^6 \pm 8,6 \times 10^6$	$244 \pm 190$	$\rightarrow 0$	$0,01 \pm 0,007$	0,993
	SIR	$641 \pm 121$	—	—	$0,196 \pm 0,002$	0,993
23 de mayo	Richards	$208 \pm 2$	$35 \pm 1$	$1,6 \pm 0,3$	$0,14 \pm 0,04$	0,998
	von Bertalanffy	$249 \pm 7$	$28 \pm 1$	$-1/3$	$0,164 \pm 0,007$	0,990
	Logístico	$211 \pm 2$	$34,0 \pm 0,2$	1	$0,176 \pm 0,006$	0,997
	Gompertz	$218 \pm 3$	$30,6 \pm 0,3$	$\rightarrow 0$	$0,111 \pm 0,006$	0,996
	SIR	$235 \pm 3$	—	—	$0,216 \pm 0,001$	0,996

de mayor  $R^2$ , debido a que posee un mayor número de parámetros de ajuste y por ende una mayor flexibilidad. Los parámetros estimados a partir del ajuste presentan, en la mayoría de los casos, un alto error estándar lo cual responde al pequeño tamaño de la muestra trabajada y a que los criterios y condiciones de muestreo fueron variando con el paso de los días. Lo anterior provocó que en esta etapa la estimación de parámetros claves como el tamaño total del brote y el máximo en la incidencia se dieran como pronósticos reservados.

La comparación entre los modelos se realizó a partir del criterio de información de Akaike, con el objetivo justificar, o no, el aumento en complejidad de modelos como el de Richards, respecto a modelos más simples. La tabla 2 muestra que el modelo con mejor bondad de ajuste del conjunto de candidatos, fue nuevamente el de Richards con un 0,911 de probabilidad. Es decir, el modelo de Richards hace mínima la pérdida de información, justificándose así la presencia de un parámetro más respecto a los demás modelos de crecimiento sigmoial y dos parámetros respecto al SIR. A pesar de los resultados obtenidos, resulta contradictorio generar un pronóstico a corto, mediado y largo plazo a partir del modelo de Richards, debido a que este indica el fin del brote en la provincia, que en ese momento presentaba altos índices de contagios localizados en la zona nordeste y en la capital provincial. Además, dicho modelo tiende a subestimar los casos acumulados a ocurrir en el período de una semana, véase la figura 3.

Por otra parte, durante el período de evaluación del pronóstico, las métricas de desempeño, ver tabla 3, indican que el comportamiento de los datos se acerca al modelo de salida más favorable, que nuevamente resultó ser el modelo de Richards. Lo anterior muestra que las medidas iniciales de

control en la provincia resultaron efectivas, evidenciándose un incremento de solo 14 casos en una semana, cuando la salida de algunos modelos mostraban escenarios más críticos con estimados en el intervalo de 67 a 111 casos acumulados. Resulta válido señalar que durante el período de pronóstico los datos siempre estuvieron dentro del cono generado por el conjunto de modelos, y más específicamente dentro de los límites establecidos por los modelos Logístico y de Richards.

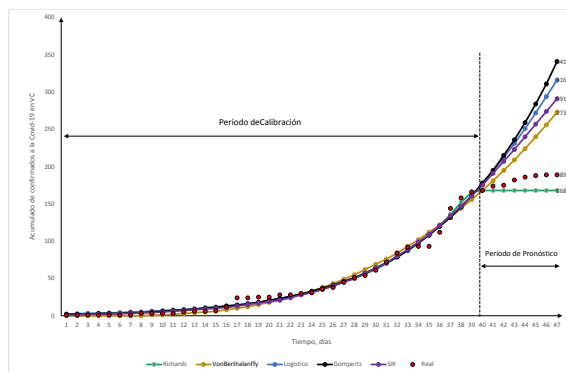
La aplicación de la ley de crecimiento exponencial en esta etapa temprana permitió estimar el número básico reproductivo en la provincia. Asumiendo que el tiempo infeccioso medio de enfermos con la Covid-19 es de 15 días y utilizando la ecuación (8) se obtuvo:  $R_0 = 3,40 \pm 0,06$ . Es decir, en los primeros momentos en la provincia cada contagiado infectaba en promedio a poco más de tres personas.

### 2.2 Etapa intermedia, del 11 de marzo al 19 de abril

El evento de transmisión local suscitado en el hogar de ancianos número 3 de Santa Clara entre el 15 y el 18 de abril, el segundo más grande en todo el país hasta el momento, caracterizó la dinámica del brote en esta etapa intermedia. Los modelos fueron ajustados a la serie de datos de contagiados acumulados hasta la fecha, de dimensión 40. Debido al trabajo con un mayor número de datos y a que estos se comportaron según la dinámica natural del brote, el ajuste realizado reportó un  $R^2$  más cercano a 1, superior o igual a 0,99 en todos los casos. Similarmente a lo ocurrido en la primera etapa, el modelo de Richards presentó el mayor índice de determinación, igual a 0,994. En cuanto a la comparación de los modelos mediante el criterio de información de Akaike, ver tabla 2, el modelo de Richards fue el de mejor bondad de ajuste con una probabilidad de 0,727. Debido a que el modelo de Richards

**Tabla 2.** Índices relacionados con el criterio de información de Akaike obtenidos a partir del ajuste de los modelos a los datos de contagiados acumulados hasta los días 29 de marzo, 19 de abril y 23 de marzo

Fecha	Modelos	AIC	$\Delta AIC$	$w_i$
29 de marzo	Exponencial	98,9969	13,8131	0,001
	Richards	85,1838	0	0,911
	von Bertalanffy	94,4039	9,2201	0,009
	Logístico	90,7737	5,5899	0,056
	Gompertz	92,5345	7,3507	0,023
	SIR	105,719	20,5352	$3,166 \times 10^{-5}$
19 de abril	Richards	251,981	0	0,727
	von Bertalanffy	273,904	21,923	$1,262 \times 10^{-5}$
	Logístico	256,694	4,713	0,069
	Gompertz	256,101	4,12	0,092
	SIR	255,735	3,754	0,111
	23 de mayo	Richards	509,645	0
von Bertalanffy		612,293	102,648	$3,862 \times 10^{-23}$
Logístico		511,87	2,225	0,247
Gompertz		548,571	38,926	$2,654 \times 10^{-9}$
SIR		544,126	34,481	$2,450 \times 10^{-8}$



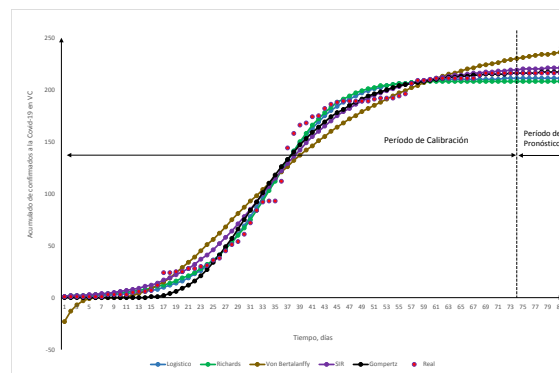
**Figura 4.** Ajuste de los modelos a la serie de datos de contagiados acumulados hasta el día 19 de abril.

mostraba escenarios de culminación del brote, al igual que en la primera etapa, y que la provincia se encontraba en medio de un alza de contagios, se consideraron los pronósticos generados a partir de las salidas de los modelos SIR y Logístico, estimándose un tamaño total del brote en el intervalo de 520 a 1385 casos, y la ocurrencia del máximo de incidencia cerca del día 50 a partir de comenzado el brote, es decir el 29 de abril.

Analizando el período de pronóstico, se observa que el comportamiento real de los datos sigue un patrón cercano al modelo de Richards, lo que se corrobora con los valores de las métricas de desempeño mostradas en la tabla 3. Una vez más prevalece el criterio de que las medidas de control de la epidemia en la provincia resultaron sumamente efectivas ya

que los resultados evidencian un crecimiento de solamente 21 casos en el acumulado, mientras el cono de pronósticos predecía que el total de casos confirmados podría haber llegado hasta alrededor de los 340 casos. Similar a la etapa anterior, los datos reales estuvieron siempre dentro del cono de pronóstico generado por todos los modelos; aunque en esta segunda etapa el cono fue más estrecho cubriendo los datos y estuvo delimitado por el modelo de Richards (límite inferior) y el modelo de von Bertalanffy (límite superior).

### 2.3 Etapa final, del 11 de marzo al 23 de mayo



**Figura 5.** Ajuste de los modelos a la serie de datos de contagiados acumulados hasta el día 23 de mayo.

Hasta esta fecha la serie de datos del acumulado de confirmados a la Covid-19 en VC constaba con 74 puntos, evidenciando una vez más el mejoramiento de los indicadores

**Tabla 3.** Métricas para evaluar los pronósticos realizados mediante los diferentes modelos para la semana siguiente a los días 29 de marzo, 19 de abril y 23 de marzo

Fecha	Modelos	RMSE	MAE	MAPE
29 de marzo	Exponencial	23,42	19,71	0,60
	Richards	7,26	5,86	0,17
	von Bertalanffy	25,22	22,57	0,70
	Logístico	8,38	8,14	0,27
	Gompertz	42,68	37,00	1,13
	SIR	16,72	14,00	0,42
19 de abril	Richards	16,42	15,29	0,08
	von Bertalanffy	49,09	42,14	0,23
	Logístico	78,02	69,57	0,37
	Gompertz	90,62	79,71	0,43
	SIR	63,58	57,14	0,31
23 de mayo	Richards	8,00	8,00	0,04
	von Bertalanffy	17,64	17,57	0,008
	Logístico	5,00	5,00	0,02
	Gompertz	0,76	0,57	0,00
	SIR	4,46	4,43	0,02

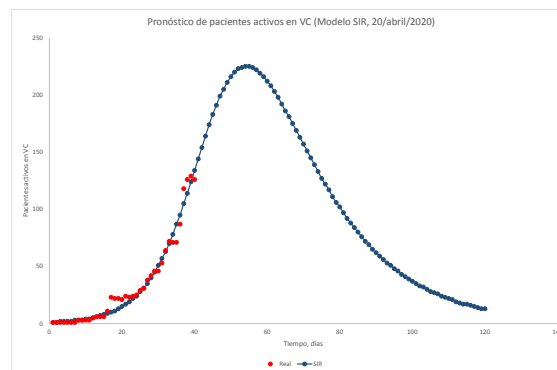
de ajuste con el tiempo. Similar a la etapa anterior, todos los modelos mostraron un  $R^2$ , superior a 0,99, como se observa en la tabla 1, y una vez más el modelo de Richards, resultó ser el más notable con  $R^2$  igual a 0,998, con un ajuste casi perfecto, reflejando excelentemente una dinámica de la enfermedad con un solo brote. También, considerando el criterio de información de Akaike, ver tabla 2, el modelo de Richards fue el de mejor bondad de ajuste con una probabilidad de 0,752. En la figura 3 se observa claramente el patrón de crecimiento sigmoïdal de todos los modelos incluidos en el análisis, reflejando a su vez un comportamiento de meseta característico del final de una epidemia de una sola oleada. A pesar de que el evento del poligráfico ?Enrique Núñez Rodríguez? vinculó a más de 700 personas, solo generó 14 nuevos casos confirmados sin afectar esta tendencia. Además, se aprecia que la variabilidad entre los modelos es similar, tanto en el período de calibración como en el período de prueba, y considerablemente menor en el período de pronóstico con relación a la etapa anterior. La incidencia de los casos confirmados a la Covid-19 en VC alcanzó su valor máximo el día 37 de la epidemia (16 de abril), en correspondencia con el evento registrado en el hogar de ancianos. Como se observa en la tabla 1, la mejor estimación para este parámetro se obtuvo por el modelo de Richards ( $t_i = 35 \pm 1$ ).

El valor real del tamaño total del brote en VC fue de 216 casos, a partir del último reporte de nuevos casos confirmados en la provincia el 18 de mayo de 2020. Analizando el período de pronóstico se observa que el comportamiento real de los datos se ajustó más a la trayectoria seguida por el modelo de Gompertz, lo que se evidencia con los valores de las métricas de rendimiento mostradas en la tabla 3. Este resultado se corrobora en la tabla 1, donde este modelo estima los valores

para el parámetro  $K$  en el intervalo de 215 a 221.

### 2.4 Sobre pronósticos del modelo SIR

El modelo SIR también fue empleado en el ajuste de la serie de datos de contagiados activos. Mediante este procedimiento se mantuvo informado a las autoridades del CDP sobre la ocurrencia del pico de casos activos y sobre el impacto de las medidas adoptadas, a partir del cálculo del número de reproducción efectivo.

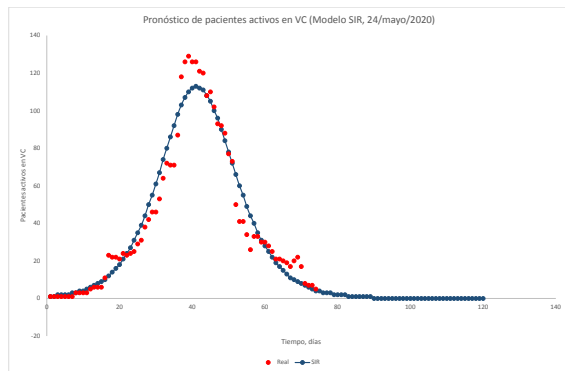


**Figura 6.** Representación gráfica del ajuste del modelo SIR a los datos de confirmados activos en la provincia de Villa Clara al cierre del 19 de abril.

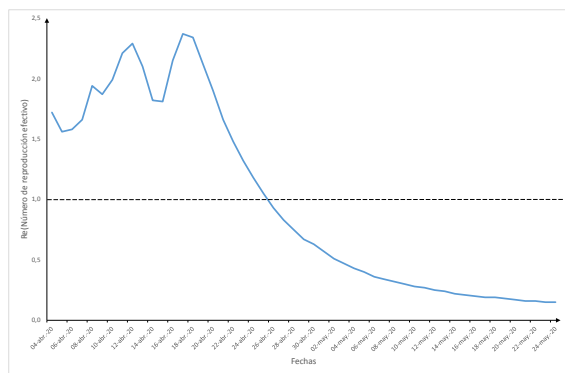
En la figura 6 se muestra el ajuste del modelo SIR a los datos de activos reportados al cierre del 19 de abril, correspondientes a la segunda etapa del análisis anterior. El coeficiente de determinación del ajuste resultó  $R^2 = 0,987$ . Este día fue



el primero en el cual se registró un mayor número de casos removidos (altas más fallecidos) que nuevos contagios, es decir es el primer día en que los casos activos comienzan a descender. Sin embargo, la salida del modelo SIR alertó sobre la ocurrencia del pico, en la provincia, sobre los días 3 y 4 de mayo con un total de 225 casos activos para este momento. A partir del 19 de abril la dinámica de casos activos continuó en descenso, constatándose la ocurrencia real del pico el 18 de abril con 129 contagiados activos. En la figura 7 se presenta el ajuste del modelo SIR a los datos de activos reportados al cierre del 23 de mayo, con índice de determinación  $R^2 = 0,977$ . El modelo estima la ocurrencia del pico el 20 de abril, dos días después del real, y 113 casos activos para esta fecha, 13 menos que el valor real. Además, predice llegar a cero casos activos el 8 de junio.



**Figura 7.** Representación gráfica del ajuste del modelo SIR a los datos de confirmados activos en la provincia de Villa Clara al cierre del 23 de mayo.



**Figura 8.** Comportamiento del número de reproducción efectivo en Villa Clara en la etapa del 4 de abril al 24 de mayo.

### 2.4.1 Sobre el número de reproducción efectivo

La figura 8 muestra el comportamiento del número de reproducción efectivo con el tiempo, en el período del 4 de abril al 24 de mayo. Puede observarse que el valor máximo de  $R_e(t)$  se alcanzó el 17 de abril, asociado al evento de transmisión local de mayor connotación en la provincia, suscitado en el hogar de anciano número 3 de Santa Clara. A partir de este día  $R_e(t)$  comenzó a disminuir gradualmente como reflejo del impacto de las medidas de control adoptadas por las autoridades y del comportamiento social. El 26 de abril el número de reproducción efectivo disminuyó por debajo de la unidad, comenzando así la etapa de control del brote epidémico.

## 3. Conclusiones

El estudio de la dinámica de transmisión del SARS CoV-2 en Villa Clara, mediante diferentes modelos matemáticos reportados en la literatura, resultó de gran utilidad para la toma de decisiones en el control adecuado sobre el desarrollo de la epidemia en la provincia. La evaluación de los pronósticos obtenidos por los diferentes modelos empleados, en las tres etapas identificadas con esta finalidad, muestra que los datos reales estuvieron siempre dentro del cono de pronóstico conformado a partir de éstos, con un comportamiento más cercano al modelo de Richards corroborado por las métricas de desempeño utilizadas.

El modelo SIR empleado en el ajuste de la serie de datos de contagiados activos, y utilizado a partir de la segunda etapa, estima la ocurrencia del pico el 20 de abril con 113 casos activos y predice llegar a cero casos activos el 8 de junio. Es también de mucha utilidad mostrar el comportamiento del número de reproducción efectivo con el tiempo, cuya disminución gradual refleja el impacto favorable de las medidas de control adoptadas, con valores por debajo de la unidad a partir del 26 de abril indicando el inicio de la etapa de control del brote epidémico. El pico real de la enfermedad en el territorio con 129 casos confirmados de COVID-19 ocurre el 18 de abril, y dos meses después: el 15 de junio, correspondiente al día 97 del brote epidémico, Villa Clara ya alcanza 28 días sin nuevos casos confirmados de transmisión local.

## Referencias

- [1] Burnham, Kenneth P. y David R. Anderson: *Multimodel Inference*. Sociological Methods & Research, 33(2):261–304, nov 2004, ISSN 0049-1241. <http://journals.sagepub.com/doi/10.1177/0049124104268644>.
- [2] Chen, Tian Mu, Jia Rui, Qiu Peng Wang, Ze Yu Zhao, Jing An Cui y Ling Yin: *A mathematical model for simulating the phase-based transmissibility of a novel coronavirus*. Infectious Diseases of Poverty, 9(1):1–8, 2020, ISSN 20499957.

- [3] Chowell, Gerardo: *Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts*. Infectious Disease Modelling, 2(3):379–398, aug 2017, ISSN 24680427.
- [4] Chowell, Gerardo, Lisa Sattenspiel, Shweta Bansal y Cé-cile Viboud: *Mathematical models to characterize early epidemic growth: A review*, sep 2016. ISSN 15710645.
- [5] Danchin, Antoine, Tuen Wai Patrick Ng y Gabriel Turinici: *A new transmission route for the propagation of the SARS-CoV-2 coronavirus*. medRxiv, página 2020.02.14.20022939, feb 2020. <https://doi.org/10.1101/2020.02.14.20022939>.
- [6] Díaz-Canel Bermúdez, M y J Núñez Jover: *Gestión gubernamental y ciencia cubana en el enfrentamiento a la COVID-19*. Anales De La Academia De Ciencias De Cuba, 10(2):e881, 2020.
- [7] Gompertz, Benjamin: *On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies*. Philosophical Transactions of the Royal Society of London, 115:513–583, dec 1825, ISSN 0261-0523. <https://royalsocietypublishing.org/doi/10.1098/rstl.1825.0026>.
- [8] Hsieh, Ying Hen, David N. Fisman y Jianhong Wu: *On epidemic modeling in real time: An application to the 2009 Novel A (H1N1) influenza outbreak in Canada*. BMC Research Notes, 3(1):283, 2010, ISSN 17560500. <http://www.biomedcentral.com/1756-0500/3/283>.
- [9] Huang, Norden E. y Fangli Qiao: *A data driven time-dependent transmission rate for tracking an epidemic: a case study of 2019-nCoV*, mar 2020. ISSN 20959281. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7128746/>.
- [10] Hurvich, Clifford M., Jeffrey S. Simonoff y Chih Ling Tsai: *Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 60(2):271–293, jan 1998, ISSN 13697412. <http://doi.wiley.com/10.1111/1467-9868.00125>.
- [11] Kermack, W O y A G Mckendrick: *A contribution to the mathematical theory of epidemics*. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 115(772):700–721, aug 1927, ISSN 0950-1207. <https://royalsocietypublishing.org/doi/10.1098/rspa.1927.0118>.
- [12] Makeham, W. M.: *On the Integral of Gompertz's Function for expressing the Values of Sums depending upon the contingency of life*. Journal of the Institute of Actuaries and Assurance Magazine, 17(5):305–327, apr 1873, ISSN 2046-1674. </core/journals/journal-of-the-institute-of-actuaries/article/on-the-integral-of-gompertz-s-function/2C1E0D5D0AE6F5054D25AF6D4F6520C1>.
- [13] Miranda, L. C.M. y C. A.S. Lima: *On the logistic modeling and forecasting of evolutionary processes: Application to human population dynamics*. Technological Forecasting and Social Change, 77(5):699–711, jun 2010, ISSN 00401625.
- [14] Montesinos-López, Osva Antonio y Carlos Moisés Hernández-Suárez: *Modelos matemáticos para enfermedades infecciosas*, 2007. ISSN 00363634.
- [15] Morais, Apiano F.: *Logistic approximations used to describe new outbreaks in the 2020 COVID-19 pandemic*. 793 in Italy, 11:405, mar 2020. <http://arxiv.org/abs/2003.11149>.
- [16] Mujib, M., M. Mardiyah, Suherman, R. Rakhmawati, S. Andriani, M. Mardiyah, H. Suyitno, S. Sukestiyarno y I. Junaidi: *The Application of Differential Equation of Verhulst Population Model on Estimation of Bandar Lampung Population*. Journal of Physics: Conference Series, 1155(1), 2019, ISSN 17426596.
- [17] Organización Mundial de la Salud (OMS): *Declaración sobre la segunda reunión del Comité de Emergencias del Reglamento Sanitario Internacional (2005) acerca del brote del nuevo coronavirus (2019-nCoV)*. Informe técnico, 2020.
- [18] Pell, Bruce, Yang Kuang, Cecile Viboud y Gerardo Chowell: *Using phenomenological models for forecasting the 2015 Ebola challenge*. Epidemics, 22:62–70, mar 2018, ISSN 18780067. <https://linkinghub.elsevier.com/retrieve/pii/S1755436516300433>.
- [19] Putter, A.: *Studies on the physiological similarity. VI. Similarities in growth*. Pflugers Archiv für die Gesamte Physiologie des Menschen und der Tiere, 180:280, 1920.
- [20] Richards, F. J.: *A flexible growth function for empirical use*. Journal of Experimental Botany, 10(2):290–301, 1959, ISSN 00220957.
- [21] Sebrango-Rodríguez, C. R., D. A. Martínez-Bello, L. Sánchez-Valdés, P. J. Thilakarathne, E. Del Fava, P. Van Der Stuyft, A. López-Quílez y Z. Shkedy: *Real-time parameter estimation of Zika outbreaks using model averaging*. Epidemiology and Infection, 145(11):2313–2323, aug 2017, ISSN 14694409. <https://doi.org/10.1017/S0950268817001078>.

- [22] Tjørve, Kathleen M.C. y Even Tjørve: *The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family*. PLoS ONE, 12(6), 2017, ISSN 19326203.
- [23] Torrealba-Rodríguez, O., R. A. Conde-Gutiérrez y A. L. Hernández-Javier: *Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models*. Chaos, Solitons and Fractals, 138:109946, sep 2020, ISSN 09600779.
- [24] Verhulst, P F: *Recherches mathématiques sur la loi d'accroissement de la population*. Nouveaux mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles, página 60, 1844.
- [25] Vynnycky, E y R White: *An Introduction to Infectious Disease Modelling*. OUP Oxford, 2010, ISBN 9780198565765. <https://books.google.com/cu/books?id=QNfqaXWL94IC>.
- [26] Wang, Xiang Sheng, Jianhong Wu y Yong Yang: *Richards model revisited: Validation by and application to infection dynamics*. Journal of Theoretical Biology, 313:12–19, 2012, ISSN 00225193. <http://dx.doi.org/10.1016/j.jtbi.2012.07.024>.
- [27] Wu, Ke, Didier Darcet, Qian Wang y Didier Sornette: *Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world*. mar 2020. <http://arxiv.org/abs/2003.05681>.
- [28] Zhou, Xiang, Na Hong, Yingying Ma, Jie He, Huizhen Jiang, Chun Liu, Guangliang Shan, Longxiang Su, Weiguo Zhu y Yun Long: *Forecasting the Worldwide Spread of COVID-19 based on Logistic Model and SEIR Model*. medRxiv, página 2020.03.26.20044289, apr 2020. <http://medrxiv.org/lookup/doi/10.1101/2020.03.26.20044289>.