

Modelos fenomenológicos aplicados al estudio de la COVID-19 en Cuba

Phenomenological models applied to the study of COVID-19 in Cuba

Jose Alejandro Mesejo-Chiong^{1*}, Angela Mireya León-Mecías¹

Resumen El reciente brote epidémico de la COVID-19 en el planeta ha resultado en una seria amenaza al desarrollo de la especie humana. Por ello se impone estudiar con todos los medios a nuestro alcance la evolución de esta enfermedad. En el presente trabajo se aplican modelos llamados fenomenológicos al estudio de la COVID-19 en Cuba. Como fuente de datos solo se requiere el reporte diario que efectúan las autoridades sanitarias de nuestro país desde el inicio de la epidemia. Se demuestra que los modelos fenomenológicos tienen gran valor para realizar pronósticos que pueden guiar las intervenciones que los sistemas nacionales de salud realizan para contener la expansión de la COVID-19.

Abstract The recent epidemic outbreak of COVID-19 on the planet has resulted in a serious threat to the development of the human species. That is why it is necessary to study, with all the means at our disposal, the evolution of this disease. In the present work we apply phenomenological models to the study of COVID-19 in Cuba. Only the public daily report, made by the health authorities of our country since the beginning of the epidemic, is used as data source. It is shown that phenomenological models have great forecasting value and its prognoses can guide the interventions carry out by national health systems to contain the expansion of COVID-19.

Palabras Clave

COVID-19, Modelos fenomenológicos, Crecimiento de epidemias

Keywords

COVID-19, Phenomenological models, Epidemic Growth

¹ Departamento de Matemática Aplicada, Facultad de Matemática y Computación, Universidad de La Habana, La Habana, Cuba, mesejo@matcom.uh.cu, angela@matcom.uh.cu

*Autor para Correspondencia, Corresponding Author

Introducción

El reciente brote epidémico de la COVID-19 causado por el virus SARS-CoV-2 demuestra que las epidemias siguen representando una seria amenaza para los seres humanos. A pesar del conocimiento acumulado en el estudio de las enfermedades infecciosas y de los tratamientos médicos actuales la COVID-19 continua su expansión mundial con un gran número de muertes y su evolución futura es aún incierta. Por ello el estudio de esta epidemia por todas las ramas de la ciencia posibles es de suma importancia.

Uno de los mayores retos está en predecir tendencias de comportamiento futuro para preparar los sistemas de salud al enfrentamiento a la epidemia. Un dato fundamental es el número de nuevos infectados por día, establecer cotas para este valor es una ayuda invaluable para la preparación y ubicación de los recursos médicos. A partir de este dato se puede deducir la duración del brote epidémico y el día en que la serie temporal de nuevos casos comenzará a demostrar una tenden-

cia decreciente. Un pronóstico acertado de estas magnitudes permite la toma de decisiones correctas.

La modelación matemática de epidemias es un tema en el cual el número de publicaciones es sumamente amplio, vea [5], [11], [12], [15], [16], [27] y [29] por solo citar algunos de los muchos textos existentes. Numerosos modelos deterministas o estocásticos se emplean para simular el desarrollo de una epidemia. Sin embargo, elegir el modelo adecuado no es una tarea fácil y la estimación eficiente de los parámetros de este modelo tampoco lo es. Se precisa una evaluación racional entre muchos posibles candidatos (vea [8]). Esta situación es particularmente compleja cuando se trata de una pandemia como la COVID-19 cuyo comportamiento presenta aún aspectos desconocidos para la ciencia.

Existen diferentes categorías de modelos matemáticos empleados en el estudio temprano de epidemias [9]. Los modelos deterministas son aquellos que se expresan mediante ecuaciones diferenciales ordinarias. A esta clase pertenecen los

clásicos y bien estudiados modelos de compartimentos SIS, SIR, SEIR y sus numerosas variantes. Otra clase la constituyen los llamados modelos fenomenológicos. Estos modelos se emplean en diferentes fenómenos como crecimiento de poblaciones, tumores, economía y epidemias, cuando se asume que la tasa de crecimiento puede ser representada mediante una única ecuación diferencial ordinaria. Entre estos podemos citar, el modelo de crecimiento exponencial (de Malthus) [21], el logístico [25], el de Richards [18] y el de Gompertz [24], así como sus variantes generalizadas, entre otros. Estos son modelos simples que hacen énfasis en incorporar observaciones empíricas y permiten reproducir características esenciales del fenómeno bajo estudio.

Dada la novedad de la COVID-19, los datos disponibles sobre su evolución son aún limitados, sobre todo si se compara con epidemias como la Gripe y otras que llevan décadas azotando a la humanidad. Los mecanismos de propagación y contagio de la COVID-19 no son totalmente conocidos (vea [1]) y las medidas de contención aplicadas están siendo evaluadas sobre bases empíricas ([3], [10], [17]). Por ello para este estudio, hemos elegido los modelos fenomenológicos no lineales simples mencionados anteriormente. El objetivo es obtener una predicción rápida a corto y mediano plazo de la cantidad de casos acumulados posibles, ayudando así al planeamiento de las medidas de contención necesarias. Este proceder se ha empleado con éxito en otras epidemias (vea [2], [13], [14], [20], [26] y [31]) y se aplica a la actual, causada por el SARS Cov 2 ([6], [7], [19], [23], [28], [30]).

En [31] se analizan datos de la epidemia del ZIKA en 2015-2016, en 8 estados de Brasil, mediante tres modelos: el logístico, el de Richards y el de Gompertz. Los tres modelos se presentan dependientes de tres parámetros fundamentales: el mayor número de casos acumulados posibles K , la tasa de crecimiento γ y el punto de inflexión τ_p . En el trabajo se demuestra que la obtención de predicciones en tiempo real confiables sobre la dimensión final de la epidemia son posibles con estos modelos.

En [30] se realiza una investigación sistemática sobre el poder de predicción de 8 funciones empíricas, 4 métodos de inferencia estadística, y 5 modelos dinámicos ampliamente usados en la literatura. Se utilizan los datos públicos del número de casos de la COVID-19 en nueve provincias Chinas durante la primavera de 2020. En este trabajo se destaca la importancia de tener un balance adecuado entre la complejidad del modelo, la precisión, el sobreajuste, el subajuste, la robustez y la sensibilidad. Se plantea que la curva sigmoidea típica de la cantidad de infectados acumulados se puede ajustar exitosamente mediante las funciones logística, de Richards y de Gompertz. En este trabajo se plantea el papel crucial que juega para el pronóstico el punto de inflexión; antes de la ocurrencia del mismo ninguno de los modelos considerados logró una predicción confiable. Se plantea además que la función logística subestima la dimensión final de la epidemia y la función de Gompertz la sobrestima en todos los casos estudiados. Es esta una afirmación que trataremos de comprobar con los

datos de la COVID-19 en Cuba.

En [4] se usan las funciones logística y de Gompertz para analizar la evolución temporal de la COVID-19 en diferentes países. Se propone además una generalización para la ley de Gompertz. En este trabajo se concluye que en China las fuertes medidas de restricción tomadas por el gobierno hicieron posible que la propagación de la enfermedad alcanzara rápidamente su punto de saturación y vaticinaron un nuevo rebrote de la epidemia en Singapur, a partir del análisis de los datos de ese país. En [2] se realiza un análisis comparativo entre el modelo de crecimiento logístico, el modelo de Gompertz y sus respectivas generalizaciones, usando datos de 37 brotes de diferentes epidemias (ébola, colera, influenza, entre otras) en diferentes regiones del mundo, con el resultado de que el modelo logístico generalizado es superior para el ajuste de datos a la evolución temporal de dichas epidemias. En [28] se obtienen cotas superiores e inferiores para el comportamiento de la COVID-19 aplicando cuatro modelos de crecimiento: el logístico, el logístico generalizado, el de crecimiento generalizado y el de Richards generalizado, al ajustar los datos de infectados reportados entre el 10 de enero y el 19 de marzo de 2020 en diferentes escenarios: toda China, 29 provincias de China y algunos países de Europa fuertemente afectados por la epidemia.

En la presente contribución se exploran los planteamientos de los trabajos citados previamente, con datos del número de casos detectados de la COVID-19 en Cuba. Nos proponemos también explorar el valor predictivo de los diferentes modelos y establecer una metodología que permita hallar cotas de la evolución temporal de la COVID-19. El trabajo está estructurado de la siguiente forma: en la sección siguiente se presentan los modelos fenomenológicos que serán empleados con sus respectivas soluciones, en la sección 2 se exponen los resultados obtenidos por el ajuste mediante mínimos cuadrados no lineales, de cada uno de los modelos a datos de Cuba y se señalan detalles de implementación relevantes, por último en la sección 3 se discuten los resultados obtenidos y se comentan las conclusiones más relevantes.

1. Modelos fenomenológicos empleados

Los modelos fenomenológicos, como se ha señalado, son especialmente adecuados cuando existe incertidumbre sobre las características epidemiológicas de la enfermedad infecciosa bajo estudio [9]. Esto es lo que ocurre con la COVID-19 por lo cual estos modelos constituyen un punto de partida, complementario a los modelos de compartimentos, para pronósticos a corto plazo. También pueden servir para una estimación temprana de la dimensión que puede alcanzar la epidemia a corto, mediano y largo plazo, aunque no son capaces de vaticinar brotes consecutivos.

Los modelos que estudiaremos se plantean mediante un problema de valor inicial (PVI) del tipo

$$\frac{dC}{dt} = f(C, \mathbf{p}), \quad C(0) = C_0, \quad (1)$$

donde $C(t)$ representa la cantidad de casos acumulados reportados hasta el instante (tiempo) t . La ecuación diferencial ordinaria en (1) es autónoma y la función f depende de un vector de parámetros del respectivo modelo que se ha denotado con \mathbf{p} . La dimensión de \mathbf{p} caracteriza la complejidad del modelo.

1.1 Modelo de Crecimiento Exponencial (MCE)

En la fase inicial de la expansión de una enfermedad infecciosa el número acumulado de nuevos casos $C(t)$, donde t tiene unidades de tiempo, demuestra un crecimiento exponencial. Esta observación empírica se expresa mediante una tasa de crecimiento proporcional al número de infectados, aseveración que es cierta en la etapa inicial de una epidemia, cuando aún no se toman medidas que permitan frenar el contagio. El modelo fenomenológico más simple para esta situación, que en el contexto de crecimiento de poblaciones se conoce como de Malthus [21], está dado por el PVI

$$\frac{dC}{dt} = \gamma C, \quad C(0) = C_0, \quad (2)$$

cuya solución analítica es

$$C(t) = C_0 \exp(\gamma t). \quad (3)$$

Tenemos entonces que el MCE presenta como parámetros a la tasa de crecimiento $\gamma > 0$ y a la cantidad inicial de individuos contagiados $C_0 > 0$. Dado que γ es positiva este modelo pronostica un crecimiento del número de infectados que solo está limitado por el número total de individuos existentes en la población. El vector de parámetros es

$$\mathbf{p}_{MCE} = (\gamma, C_0) \quad (4)$$

y evidentemente la curva dada por (3) no presenta puntos de inflexión.

1.2 Modelo de Crecimiento Exponencial Generalizado (MCEG)

Al igual que el MCE, el MCEG [26] es aplicable en la fase inicial de la epidemia. Está representado por el PVI

$$\frac{dC}{dt} = \gamma C^\mu, \quad C(0) = C_0, \quad (5)$$

donde $0 < \mu < 1$ caracteriza la desviación del MCEG respecto al MCE. Este tipo de crecimiento se llama subexponencial. La solución general de (5) es

$$C(t) = (\tilde{\gamma}t + \tilde{C}_0)^{1/\bar{\mu}} \quad (6)$$

donde

$$\tilde{\gamma} = \gamma \bar{\mu}, \quad \bar{\mu} = 1 - \mu, \quad \tilde{C}_0 = C_0 \exp(-\mu C_0).$$

El vector de parámetros es

$$\mathbf{p}_{MCEG} = (\tilde{\gamma}, \bar{\mu}, \tilde{C}_0) \quad (7)$$

y a partir de este se pueden calcular los parámetros que aparecen en (5). Este modelo también pronostica un crecimiento ilimitado del número de individuos infectados.

1.3 Modelo de Crecimiento Logístico (MCL)

El modelo de crecimiento logístico [25] está dado por el PVI

$$\frac{dC}{dt} = \gamma C \left(1 - \frac{C}{K}\right), \quad C(0) = C_0. \quad (8)$$

En (8) aparecen la tasa de crecimiento γ y la llamada capacidad de carga K que determina el mayor número de individuos que podrán ser contagiados. La solución analítica de (8) es

$$C(t) = \frac{K}{1 + (K/C_0 - 1) \exp(-\gamma t)},$$

y se puede expresar incluyendo el punto de inflexión τ_p de la forma

$$C(t) = \frac{K}{1 + \exp(-\gamma(t - \tau_p))}, \quad (9)$$

considerando que

$$\tau_p = \frac{1}{\gamma} \ln \left(\frac{K}{C_0} - 1 \right). \quad (10)$$

El punto de inflexión τ_p determina el momento en el cual el número de nuevos infectados diarios empieza a decrecer o análogamente, el momento en el cual comienza la fase de desaceleración en el crecimiento de la cantidad de individuos infectados acumulados. El vector de parámetros para este modelo es

$$\mathbf{p}_{MCL} = (\gamma, K, \tau_p). \quad (11)$$

1.4 Modelo de Crecimiento Logístico Generalizado (MCLG)

El MCLG [8] está dado por el PVI

$$\frac{dC}{dt} = \gamma C^\mu (1 - \beta C), \quad C(0) = C_0, \quad (12)$$

donde de nuevo $\beta = K^{-1}$. El MCLG generaliza el MCL al incluir un nuevo parámetro μ con $0 < \mu < 1$, que cuantifica la desviación en el número acumulado de infectados de la tendencia exponencial inicial del MCL. Si $\mu = 1$ se tiene el MCL y dado que $\mu < 1$, el crecimiento inicial presenta un comportamiento subexponencial.

No existe solución analítica para (12), por lo cual, se requiere de métodos numéricos para hallar la solución, dado un vector de parámetros

$$\mathbf{p}_{MCLG} = (\beta, \gamma, \mu, C_0) \quad (13)$$

específico. El valor de K se obtiene nuevamente a partir de $K = \beta^{-1}$.

El punto de inflexión τ_p en este modelo puede ser calculado también con métodos numéricos. Se cumple que

$$\frac{d^2C}{dt^2} = \gamma C^\mu (\mu C^{-1} - \beta(\mu + 1)) \frac{dC}{dt},$$

por lo cual, τ_p satisface la ecuación

$$C(\tau_p) - \frac{\mu}{\beta(\mu + 1)} = 0. \quad (14)$$

Hallar una solución de (14) con $\tau_p > 0$ por métodos numéricos es sencillo. Observemos que con \mathbf{p}_{MCLG} podemos determinar el vector de parámetros

$$\bar{\mathbf{p}}_{MCLG} = (K, \gamma, \tau_p, C_0).$$

1.5 Modelo de Crecimiento de Richards (MCR)

EL MCR es una modificación propuesta por Richards en 1959 (vea [18]) al modelo logístico, como forma simple de ajustar datos empíricos en estudios botánicos. Está dado por el PVI

$$\frac{dC}{dt} = \gamma C \left(1 - \left(\frac{C}{K} \right)^\alpha \right), \quad C(0) = C_0. \quad (15)$$

Tiene la solución analítica

$$C(t) = \frac{K}{(1 + \exp(-\alpha\gamma(t - \tau_p)))^{1/\alpha}}. \quad (16)$$

Esta curva en forma de sigmoide tiene el punto de inflexión

$$\tau_p = \frac{1}{\alpha\gamma} \ln \left(\frac{C_0^{-\alpha} - K^{-\alpha}}{K^{-\alpha}} \right), \quad (17)$$

y el vector de parámetros es

$$\mathbf{p}_{MCR} = (K, \alpha, \gamma, \tau_p). \quad (18)$$

1.6 Modelo de Crecimiento de Richards Generalizado (MCRG)

El MCRG [8] se define mediante el PVI

$$\frac{dC}{dt} = \gamma C^\mu (1 - \beta C^\alpha), \quad C(0) = C_0, \quad (19)$$

con $\beta = K^{-\alpha}$.

Este modelo generaliza el MCL, el MCLG y el MCR al incluir los parámetros $0 < \alpha < 1$ y $0 < \mu < 1$. Como se señaló anteriormente, α permite considerar diferentes perfiles de crecimiento subexponencial y μ mide la desviación que experimenta la sigmoide descrita por la solución $C(t)$, respecto a la simétrica de la curva logística simple. La solución analítica no existe y por ello el PVI (19) se resuelve mediante métodos numéricos.

El punto de inflexión τ_p también se calcula mediante un método numérico. Observemos que

$$\frac{d^2C}{dt^2} = -\gamma C^{\mu-1} ((\alpha + \mu) \beta C^\alpha - \mu) \frac{dC}{dt}$$

Entonces τ_p es solución de la ecuación

$$C(\tau_p) - \left(\frac{\mu}{(\alpha + \mu) \beta} \right)^{1/\alpha} = 0.$$

El vector de parámetros es

$$\mathbf{p}_{MCRG} = (\alpha, \beta, \gamma, \mu, C_0) \quad (20)$$

y este nos permite calcular el vector

$$\bar{\mathbf{p}}_{MCRG} = (\alpha, K, \gamma, \mu, \tau_p, C_0).$$

1.7 Modelo de Crecimiento de Gompertz (MCG)

La llamada curva de Gompertz [31], expresada en una de sus variantes por $y = Ke^{-e^{-\gamma t}}$, fue por muchos años solo de interés para los actuarios. Sin embargo, posteriormente ha sido usada por varios autores como curva de crecimiento en fenómenos biológicos y económicos, [2], [24]. El modelo responde al PVI

$$\frac{dC}{dt} = \gamma C \ln \left(\frac{K}{C} \right), \quad C(0) = C_0, \quad (21)$$

con solución analítica

$$C(t) = K \exp \left(-\ln \left(\frac{K}{C_0} \right) \exp(-\gamma t) \right),$$

que puede escribirse en término del punto de inflexión τ_p como

$$C(t) = K \exp(-\exp(-\gamma(t - \tau_p))) \quad (22)$$

con

$$\tau_p = \frac{1}{\gamma} \ln \left(\ln \left(\frac{K}{C_0} \right) \right).$$

El vector de parámetros es

$$\mathbf{p}_{MCG} = (\gamma, K, \tau_p). \quad (23)$$

Sea

$$v(t) = \ln \left(\frac{K}{C} \right) = \ln(K) - \ln(C),$$

entonces, dado que $dv/dt = -(dC/dt)/C$, (21) se transforma en

$$\frac{dv}{dt} = -\gamma v, \quad v(0) = \ln(K) - \ln(C_0).$$

Se obtiene que

$$v(t) = (\ln(K) - \ln(C_0)) \exp(-\gamma t),$$

por lo cual una formulación equivalente de (21) es

$$\frac{dC}{dt} = \beta C \exp(-\gamma t), \quad C(0) = C_0, \quad (24)$$

con $\beta = \gamma(\ln(K) - \ln(C_0))$. La solución analítica de (24) es

$$C(t) = C_0 \exp \left(\frac{\beta}{\gamma} \right) \exp \left(-\frac{\beta}{\gamma} \exp(-\gamma t) \right).$$

1.8 Modelo de Crecimiento de Gompertz Generalizado (MCGG)

Tomando en cuenta la forma (24) del PVI para el MCG planteamos el PVI del MCGG [2] como

$$\frac{dC}{dt} = \beta C^\mu \exp(-\gamma t), \quad C(0) = C_0. \quad (25)$$

La solución analítica de (25) es

$$C(t) = C_0 \left(1 + \bar{\beta} (1 - \exp(-\gamma t)) \right)^{\bar{\mu}} \quad (26)$$

donde

$$\bar{\mu} = \frac{1}{1-\mu}, \quad \bar{\beta} = \frac{\beta}{\bar{\mu} C_0^{1/\bar{\mu}} \gamma}.$$

Dado que $\gamma > 0$ tenemos que

$$K = \lim_{t \rightarrow \infty} C(t) = C_0 (1 + \bar{\beta})^{\bar{\mu}}.$$

La segunda derivada de la función en (26) nos permite calcular que

$$\tau_p = -\frac{1}{\gamma} \ln \left(\frac{1 + \bar{\beta}}{\bar{\beta} \bar{\mu}} \right). \quad (27)$$

El vector de parámetros está dado por

$$\mathbf{p}_{MCGG} = (\bar{\beta}, \gamma, \bar{\mu}, \bar{C}_0). \quad (28)$$

1.9 Consideraciones para todos los modelos

Los modelos presentados en la secciones precedentes, excepto el MCE y el MCEG, permiten estimar dos parámetros de importancia que caracterizan una epidemia. Uno de ellos es el número máximo posible K de casos acumulados, que permite estimar cuan grave será la expansión de la epidemia. El otro es el punto de inflexión τ_p que presentará la serie de tiempo del número de casos acumulados. Esta serie de tiempo tiene que ser no decreciente y obligatoriamente, en un período de tiempo considerable, debe presentar una asíntota horizontal en una población finita.

Si, se permite que tanto α como μ tomen valor 1, el MCRG contiene al MCL, MCLG y MCR. No obstante, hemos decidido presentarlos por separado para exponer las soluciones analíticas (9) y (16) del MCL y el MCR respectivamente. Estas soluciones, junto a (3) del MCE, (6) del MCEG, (22) del MCG y (26) del MCGG son las que se emplean en el ajuste de datos a los respectivos modelos. Emplear soluciones analíticas en el ajuste de datos, es preferible a emplear aquellas que se obtienen mediante un método de integración numérica del PVI correspondiente. Esta última variante es la única posible para el MCRG.

2. Ajuste de los modelos a datos de Cuba

Con el objetivo de evaluar el desempeño de los modelos en el pronóstico del número acumulado de casos confirmados para la epidemia de COVID-19 se realizan varios experimentos numéricos con datos del brote epidémico de esta enfermedad. Los datos han sido obtenidos de los reportes diarios sobre casos confirmados emitidos por el Ministerio de Salud Pública de la República de Cuba del 11 de marzo de 2020 al 11 de mayo 2020.

El método de ajuste empleado será el clásico mínimos cuadrados no lineales (vea [8]). Sea $\mathcal{C} = \{C_i = C(t_i)\}_{i=0, \dots, n}$ la serie de tiempo de casos confirmados acumulados donde t es el tiempo en días. Tomemos como ejemplo el ajuste al MCL, entonces el valor óptimo de $\mathbf{p}_{MCL} = (\gamma, K, \tau_p)$ se busca como solución de

$$\min_{\mathbf{p}_{MCL}} \frac{1}{2} \sum_{i=0}^n (C_i - F(t_i, \mathbf{p}_{MCL}))^2, \quad (29)$$

donde, de acuerdo a (9),

$$F(t_i, \mathbf{p}_{MCL}) = \frac{K}{1 + \exp(-\gamma(t - \tau_p))}.$$

De acuerdo a (10) el parámetro τ_p que aparece en (29) contiene el valor de C_0 . Aunque el valor de C_0 de la serie de tiempo \mathcal{C} es conocido (en el caso de Cuba $C_0 = 3$), es preferible incluir su valor entre los parámetros a estimar mediante (29) y los problemas similares para los otros modelos empleados.

La solución de (29) y similares problemas obtenidos para el MCE, MCEG, MCLG, MCR, MCRG, MCG y MCGG se obtiene con la función `fit` perteneciente a la Curve Fitting Toolbox de Matlab. Esta función permite una estimación robusta al emplear el método de mínimos cuadrados reiterativamente ponderados [22] (iteratively reweighted least-squares, en inglés). Esta elección se fundamenta en que los valores de la serie de tiempo \mathcal{C} pueden estar sujetos a ruido y presentar valores atípicos. Los mínimos cuadrados no lineales clásicos suelen ser afectados por la presencia de valores atípicos en los datos.

Para la estimación robusta mediante la función `fit` existen dos variantes que se activan al emplear la opción '`Robust`' con uno de los valores '`Bisquare`' o '`LAR`'. En este trabajo se emplean ambas para comparar resultados y establecer el más apropiado. Para más detalles sobre la función `fit` vea la documentación de la Curve Fitting Toolbox.

Como se mencionó anteriormente los datos pueden contener ruido y valores atípicos. Esto no es inusual ya que los resultados de pruebas para detectar la COVID-19, como el PCR, realizados en un día pueden ser reportados en días siguientes. Por ello los valores de la serie de tiempo \mathcal{C} pueden presentar saltos considerables que no están relacionados con la evolución real de la epidemia y afectan la calidad de la estimación dada por (29). Esto indica que los valores en \mathcal{C} pueden ser considerados con cierta flexibilidad. Para demostrar esta afirmación y reducir la ocurrencia de grandes saltos en \mathcal{C} hemos aplicado el ajuste a una versión suavizada $\bar{\mathcal{C}}$ de los datos. Los valores en $\bar{\mathcal{C}}$ se calculan mediante media móvil con un largo de 7 días. Es decir, el valor de $\bar{C}_i \in \bar{\mathcal{C}}$ está dado por

$$\bar{C}_i = \frac{C_{i-3} + \dots + C_i + \dots + C_{i+3}}{7} \quad (30)$$

suprimiendo los valores a la izquierda de C_i que tengan subíndice menor que 0 y aquellos a la derecha con índice mayor que n . Esta variante de suavizado garantiza que $\bar{C}_n = C_n$ y además conserva las propiedades fundamentales del desarrollo de la epidemia. De aquí en adelante, llamaremos *datos suavizados* a aquellos que se obtienen aplicando el procedimiento de media móvil descrito anteriormente.

2.1 Valor predictivo de los modelos

El objetivo del presente trabajo es evaluar la capacidad de los modelos fenomenológicos presentados en la sección anterior para reproducir la evolución de la epidemia de COVID-19 en Cuba y además predecir dicha evolución. Para ello hemos

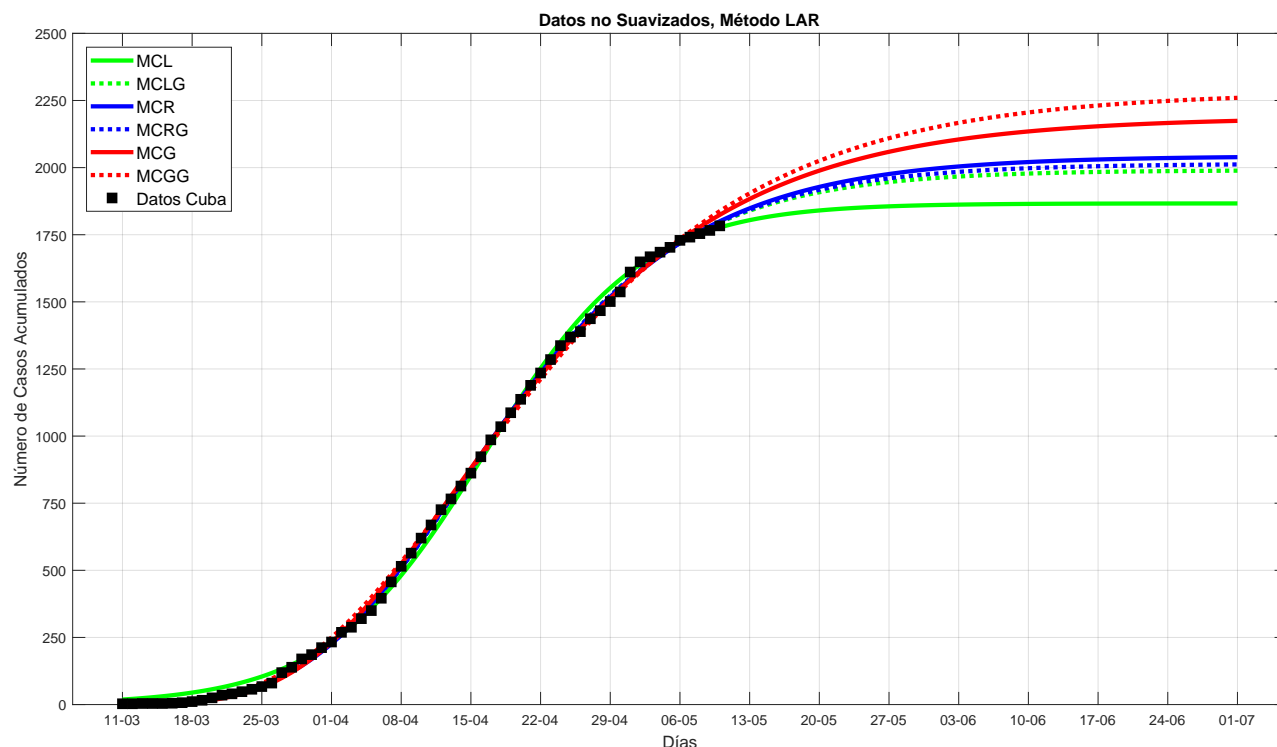


Figura 1. Resultados obtenidos del ajuste de todos los modelos empleando 61 días de datos de casos confirmados acumulados. Los resultados se obtienen empleando el método robusto LAR.

desarrollado una serie de experimentos que describimos a continuación:

1. Tomar la totalidad de la serie de tiempo \mathcal{C} de casos confirmados acumulados de la COVID-19 y realizar el ajuste de parámetros de todos los modelos a estos datos. El objetivo de este experimento radica en evaluar la capacidad de cada modelo para reproducir la curva de casos acumulados. También permite evaluar el desarrollo de la COVID-19 a largo plazo ya que los modelos MCL, MCLG, MCR, MCRG, MCG y MCGG reportan un número máximo de casos posibles. Hasta el momento de la realización de este experimento la serie de tiempo \mathcal{C} constaba de 61 datos correspondientes a los días comprendidos entre el 11 de Marzo y el 11 de Mayo de 2020.
2. Tomar una parte de la serie de tiempo \mathcal{C} y evaluar el poder predictivo de cada modelo. Para este experimento hemos tomado todos los datos de \mathcal{C} comprendidos desde su inicio hasta 3, 4, 5, 6 y 7 semanas. Es decir, se toman los primeros 21, 28 y así sucesivamente hasta 49 días de datos, se realiza el ajuste de parámetros de cada modelo a estos datos. Una vez obtenidos los parámetros se predice con cada modelo el comportamiento de la epidemia a los días no incluidos en el ajuste y se compara con los datos reales.

Los experimentos descritos en los puntos 1 y 2 arriba, se

realizan para cada uno de los ocho modelos con cada una de las variantes Bisquare y LAR (mencionadas anteriormente) que permite la función `fit` perteneciente a la Curve Fitting Toolbox de Matlab. Una forma sencilla de evaluar la calidad del ajuste está dada por el error cuadrático medio (RMSE en inglés), este error lo calcula `fit` por defecto. Se debe mencionar también que esta función halla los intervalos de confianza al 95 % para los parámetros estimados (vea la documentación).

En la figura 1 se ilustra el primer experimento, descrito en el punto 1, con datos no suavizados y el empleo del método robusto LAR. En dicha figura aparecen representados los resultados de los métodos MCL, MCLG, MCR, MCRG, MCG y MCGG. Los resultados del MCE y el MCEG no aparecen en esta ni otras figuras ya que estos modelos predicen un crecimiento exponencial sostenido que a los 61 días de epidemia en Cuba siempre sobrestima el número de casos considerablemente. Como se puede apreciar en la figura 1 todos los modelos reproducen con bastante exactitud el comportamiento de la epidemia. El RMSE del MCL es de 25,88, el del MCLG 11,93, el del MCR 12,56, el del MCRG 12,19, el del MCG 16,14 y el del MCGG 20,79. El MCL y el MCGG tienen los peores ajustes uno por debajo (subestimando) y el otro sobrestimando el número de casos posibles. El mejor ajuste se obtiene con el MCLG. De acuerdo a los métodos MCL, MCLG, MCR y MCRG y según se observa en la figura 1 a mediados de Mayo la COVID-19 entrará en una fase de estabilización donde el incremento diario del número de casos

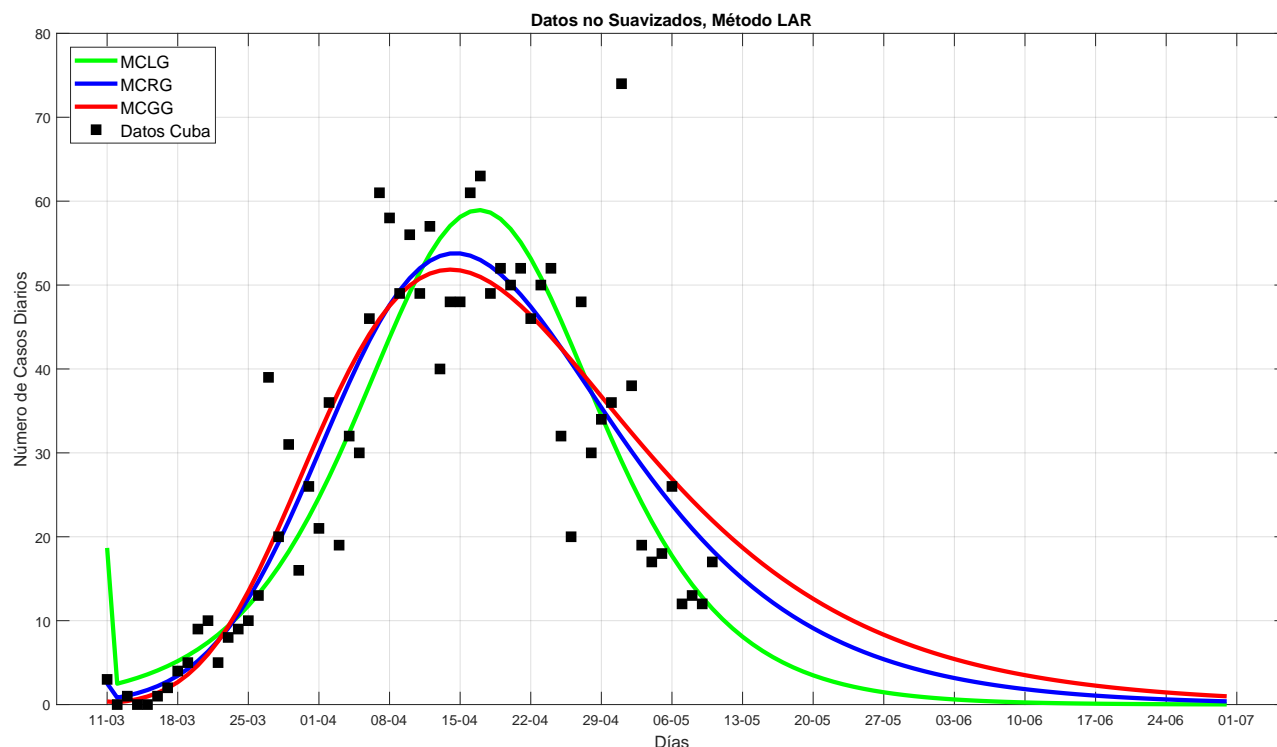


Figura 2. Resultados obtenidos del ajuste del MCL, MCR y MCG empleando 61 días de datos de casos confirmados acumulados. Los resultados se obtienen empleando el método robusto LAR y se muestran en forma de casos diarios reportados.

confirmados será pequeño. El valor máximo que predicen el MCL y el MCLG es de 1867 y 1991 casos respectivamente y los puntos de inflexión se ubican en los días 35 y 36 desde el inicio de la epidemia. El MCR y el MCRG arrojan 2043 y 2015 casos máximos respectivamente y ubican los puntos de inflexión en los días 36 y 34. Por último el MCG y el MCGG calculan 2188 y 2282 casos, ubicando ambos el punto de inflexión a los 34 días desde el inicio de la epidemia.

En resumen, con los datos de 61 días de epidemia los modelos predicen un número máximo de casos entre 1867 y 2282 con alta probabilidad de que la realidad esté entre los 2015 y 2043 casos que calculan el MCRG y el MCR. Esta cota se alcanzará rápidamente desde finales de Mayo. Todos los métodos ubican la culminación de la fase de crecimiento exponencial sobre los 34 a 36 días del 11 de Marzo, es decir del 14 al 15 de Abril de 2020.

La figura 2 muestra los mismos resultados de la figura 1 para los modelos MCL, MCR y MCG. La diferencia radica en que los resultados se reproducen como casos confirmados por día que no es más que el valor de $C_{i+1} - C_i$ con $C_i, C_{i+1} \in \mathcal{C}$. Como se aprecia en la figura 2 el MCL ubica el “pico” de la epidemia en el 17 de Abril y concede a este el mayor valor entre los 3 modelos presentados. El MCL pronostica también el descenso más rápido en el número de casos diarios. De acuerdo al MCL después del 11 de Mayo el número de casos diarios reportados debe ser menor que 10. Las fechas del MCR y MCG para este valor son el 17 y 24 de Mayo respectivamente.

La figura 3 muestra los resultados obtenidos por el MCL, el MCR y el MCG al emplear datos suavizados y el método robusto LAR. No existen diferencias significativas con respecto a los resultados mostrados en la figura 2. Solo que en la figura 3 se muestran los datos reales de Cuba en su variante suavizada. Esta forma de representación permite distinguir mejor la tendencia de la epidemia. Se puede observar que el simple suavizado de los datos permite representar con claridad el “pico” de la serie de tiempo del número de casos diarios. Volvemos a señalar que esta forma de manejar los datos afecta poco las predicciones a largo plazo, cuando el ajuste de datos se realiza contra número de casos confirmados acumulados. Es nuestra impresión que esta misma situación debe producirse si se emplean datos suavizados en el ajuste de modelos de compartimentos.

Los ajustes obtenidos mediante los métodos MCE y MCEG no aparecen en gráficas debido a que, como se mencionó, tienen crecimiento exponencial sostenido. De ser incluidos en gráficas el escalamiento del eje de las ordenadas dificultaría la visibilidad de los restantes ajustes. No obstante estos métodos tienen la propiedad de identificar con exactitud la fase de crecimiento exponencial de una epidemia. Más aún permiten constatar si las medidas de contención y prevención que se aplican son exitosas. Los parámetros para medir este desempeño son el γ del MCE y el $\bar{\mu}$ del MCEG, vea las ecuaciones (3) y (6) respectivamente. A menor valor de γ menor tasa de crecimiento exponencial, por el contrario a mayor valor de $\bar{\mu}$ menor tasa de crecimiento subexponencial. Las tablas 1 y

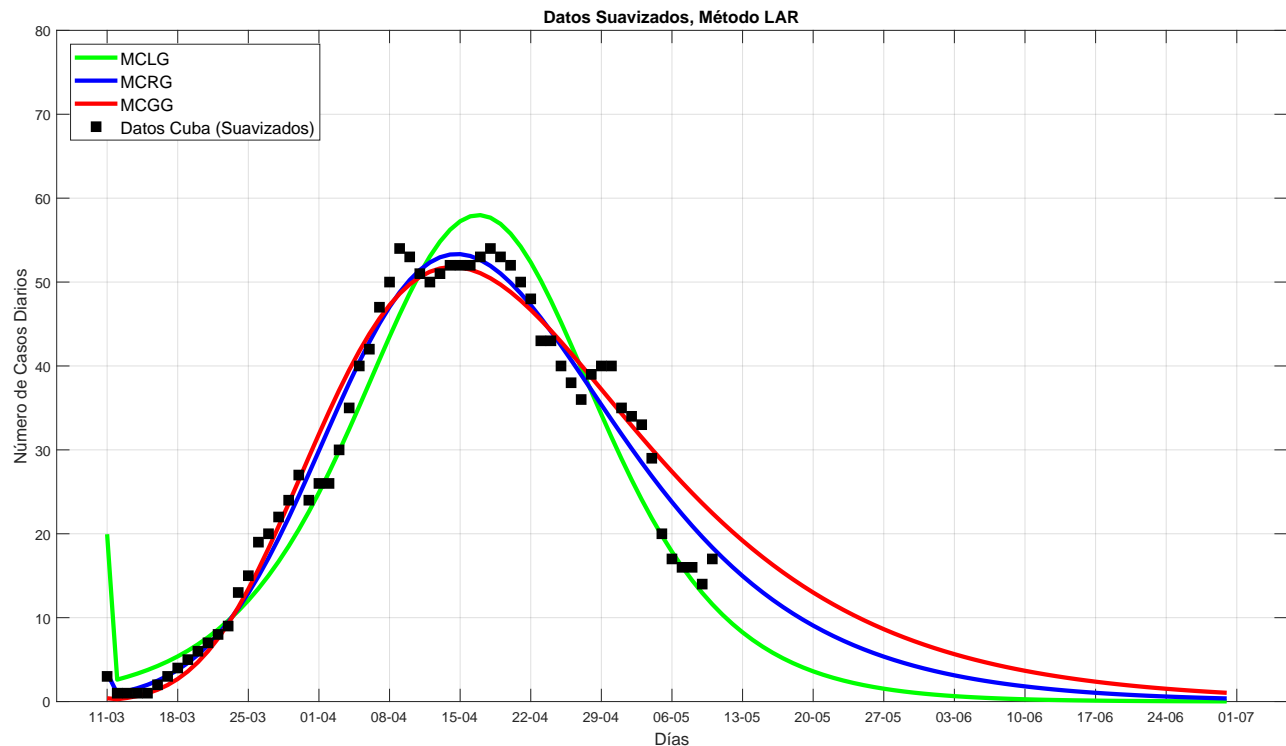


Figura 3. Resultados obtenidos del ajuste del MCL, MCR y MCG empleando 61 días de datos suavizados. Los resultados se obtienen empleando el método robusto LAR y se muestran en forma de casos diarios reportados. Los datos de Cuba aparecen en su variante suavizada.

Tabla 1. Valores de los parámetros ajustados para el MCE tomando datos parciales de 3 a 7 semanas y todos los datos disponibles (última fila).

Días	Bisquare		LAR	
	γ	C_0	γ	C_0
21	0.22417	3.04889	0.19922	4.22322
28	0.13397	12.7045	0.14568	9.16943
35	0.10897	22.3276	0.10706	24.2119
42	0.08409	42.8058	0.08786	38.206
49	0.06544	75.0492	0.07088	60.9974
61	0.04448	153.262	0.04678	139.79

2 permiten verificar estos valores para el MCE y el MCEG respectivamente, siguiendo la metodología explicada anteriormente de tomar una parte de los datos existentes y realizar los ajustes de parámetros con este subconjunto.

Como se observa en la tabla 1, a 3 semanas (21 días) de epidemia, el valor de γ era de 0,22417 y ha ido descendiendo en Cuba hasta menos de su quinta parte con el 0,04448 obtenido a 61 días de epidemia. También se aprecia en la tabla 2, que el valor de $\bar{\mu}$ asciende de 0,1222 a 0,5797. Esto refleja el éxito de las medidas de contención y prevención para la COVID-19 aplicadas en Cuba. El lector interesado podrá comprobar que tanto el descenso de γ y el incremento de $\bar{\mu}$ están temporalmente en proporción a las medidas aplicadas como

cierre de aeropuertos, restricciones al movimiento urbano e interprovincial y obligatoriedad del empleo de mascarillas faciales. En las tablas 1 y 2 se presentan los resultados con ambos métodos robustos Bisquare y LAR. El valor del parámetro C_0 en la tabla 1 depende del número de casos máximo que se presentan en los datos analizados. Aparentemente el valor de C_0 no influye en los ajustes del MCEG, ya que en todos los casos, excepto en uno, está en el valor mínimo que se le exige a la función fit para el ajuste. Nuestra suposición, que verificaremos en trabajos futuros, es que es preferible tomar $C_0 = 3$ (caso de Cuba) en la función de ajuste dada por (6).

Tabla 2. Valores de los parámetros ajustados para el MCEG tomando datos parciales de 3 a 7 semanas y todos los datos disponibles (última fila).

Días	Bisquare			LAR		
	γ	$\bar{\mu}$	C_0	γ	$\bar{\mu}$	C_0
21	0.0457	0.1222	1.05	0.1008	0.2048	1
28	0.2043	0.3069	1	0.1775	0.2871	1
35	0.2045	0.3065	1	0.2168	0.3145	1
42	0.2837	0.3536	1	0.2823	0.3536	1
49	0.5297	0.4432	1	0.4561	0.4227	1
61	1.3627	0.5797	1	1.1377	0.5531	1

El estudio del valor predictivo de los modelos se muestra en las figuras 4 a la 8. En cada una de ellas se toman varias

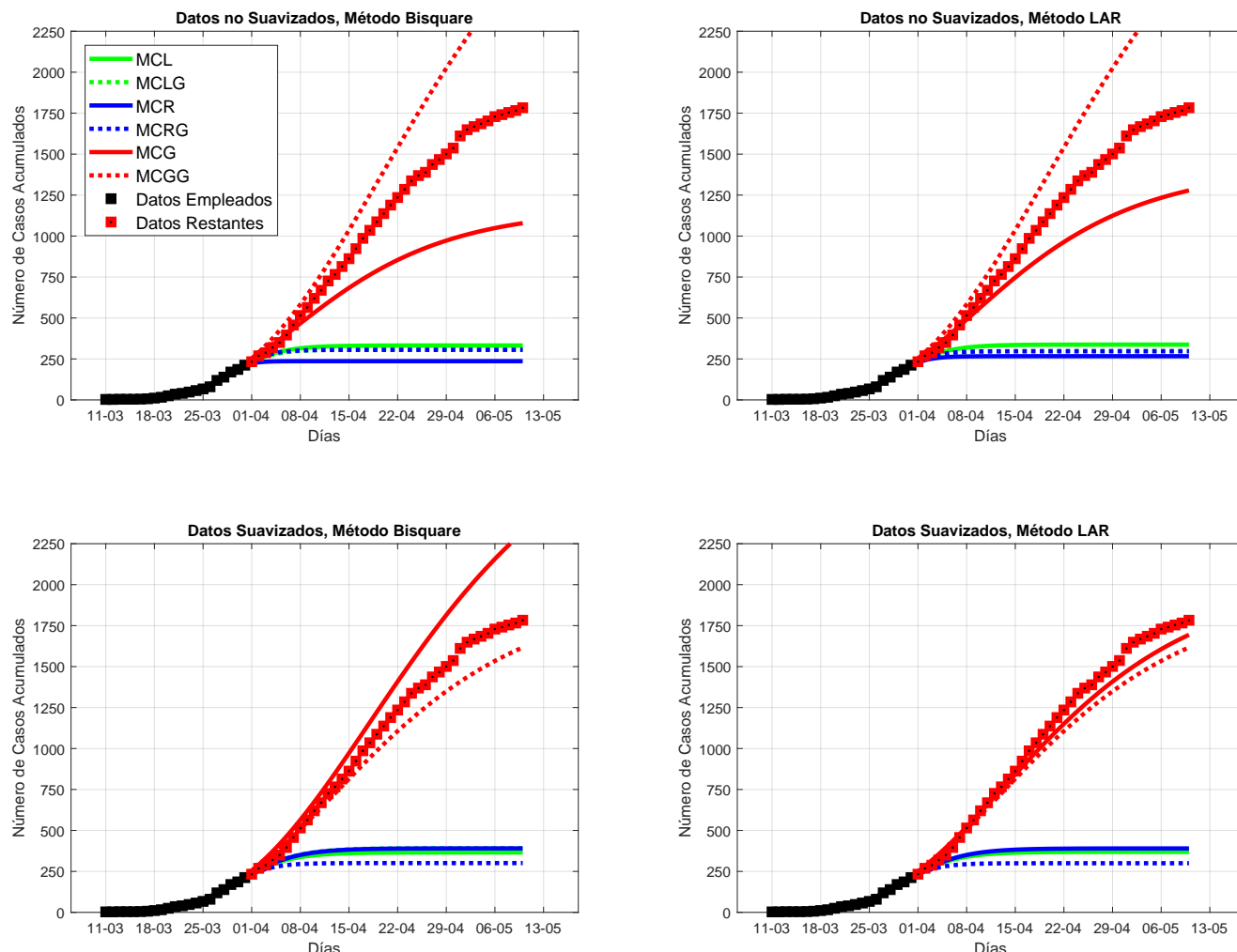


Figura 4. Resultados obtenidos empleando 3 semanas de datos (21 puntos de datos).

semanas de datos, a partir de 3, de \mathcal{C} y se realiza el ajuste de parámetros de cada modelo a esos datos. También se toman las mismas cantidades de datos de la serie suavizada \mathcal{C} , vea (30), y se realiza el ajuste con los métodos robustos Bisquare y LAR. Las figuras presentan los resultados de los cuatro ajustes de parámetros.

En la figura 4 se demuestra que pocos datos no permiten una estimación acertada a largo o mediano plazo del comportamiento de la epidemia. El único método capaz de acercarse su evaluación a los datos reales es el MCGG que sobrestima cuando se emplean los datos originales (primera fila de la figura 3). Esta figura demuestra la ventaja de usar datos suavizados cuando el conjunto de datos utilizados es pequeño (21 datos en este caso). Este comportamiento se debe a que los reportes de casos confirmados de la COVID-19 en países y regiones del planeta comienzan con varios días reportando 1 o muy pocos casos, para crecer a partir de la segunda o tercera semana. Nuestra explicación es que los sistemas de salud no comienzan un muestreo amplio de la población hasta que no

se acumula evidencia de la expansión de la COVID-19. Se debe señalar que, el proceder adoptado con datos suavizados y empleando el método LAR produce un ajuste donde los métodos MCG y MCGG permiten predecir el curso de la epidemia con bastante exactitud hasta al menos dos semanas y media en el futuro. Sin embargo, los restantes métodos producen predicciones a menos de una semana, que quedan por debajo de la evolución de la epidemia.

En cuanto aumenta la cantidad de datos el conjunto de trayectorias que predicen los modelos empieza a acotar con mayor grado de exactitud el curso de la epidemia. Esto se puede observar en las figuras de la 5 a la 8. En la figura 5 los modelos generalizados sobrestiman el curso de la epidemia y los comunes la subestiman. Sin embargo, el MCG con el método LAR para datos originales presenta un resultado muy bueno, que da un pronóstico que se extiende con exactitud por casi 4 semanas. En la figura 6 el MCR presenta un pobre desempeño si se emplean los datos originales. Por el contrario, cuando se suavizan los datos es el que mejores resultados

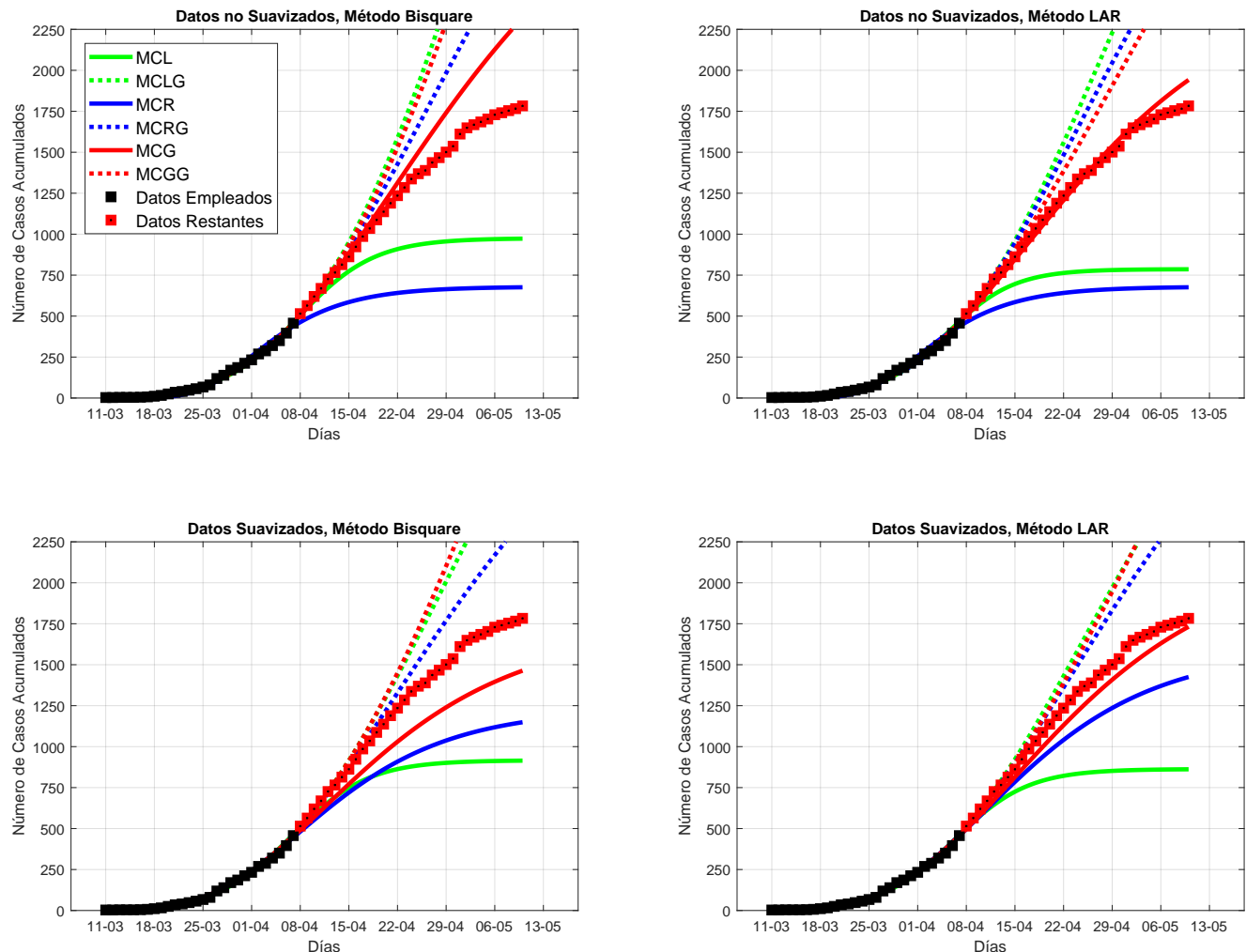


Figura 5. Resultados obtenidos empleando 4 semanas de datos (28 puntos de datos).

demuestra.

En las figuras 7 y 8 los conjuntos de trayectorias continúan proporcionando cotas cada vez más cercanas a los valores reales de casos confirmados, reportados por el MINSAP en nuestro país. Esto evidencia que a mayor disponibilidad de datos, los modelos producen curvas que se aproximan cada vez más a las reales e incrementan su valor predictivo. Es decir, cuando se toma un número de datos apreciable, en el caso de la COVID-19 mayor de 40 puntos de datos, los modelos pueden predecir con bastante exactitud el curso de la epidemia en un plazo de al menos una semana. Para hacer más confiable esta predicción, se puede tomar como cota inferior el valor de nuevos casos que proporciona el MCL y como cota superior el que proporciona el MCGG.

Las figuras 7 y 8 empiezan a mostrar la evidencia planteada por [30] de que el MCL subestima y el MCG sobrestima el curso de la epidemia. En estas figuras el MCLG apenas se aprecia ya que su trayectoria prácticamente coincide con la que describen los datos reales. El RMSE del MCLG en

la figura 7 es de 8,75 para datos no suavizados con método Bisquare, de 9,50 para estos mismos datos y método LAR, de 5,10 para datos suavizados con método Bisquare y de 4,85 para estos datos y método LAR. Estos mismos valores de RMSE del MCL para el ajuste mostrado en la figura 8 son 9,53, 9,58, 4,90 y 4,94 respectivamente.

3. Discusión y Conclusiones

Como se señala en [9], los pronósticos a corto plazo obtenidos con los modelos fenomenológicos constituyen una guía para el tipo e intensidad de las intervenciones que los sistemas de salud deban tomar en el manejo de un brote epidémico. Un mejor entendimiento de las características de una epidemia solo se puede alcanzar mediante el empleo de diferentes métodos de modelación matemática, con diferentes enfoques y el empleo de datos fidedignos sobre la enfermedad. Por ello, el estudio de la COVID-19 debe ser abordado con todas las técnicas de modelación matemática posibles.

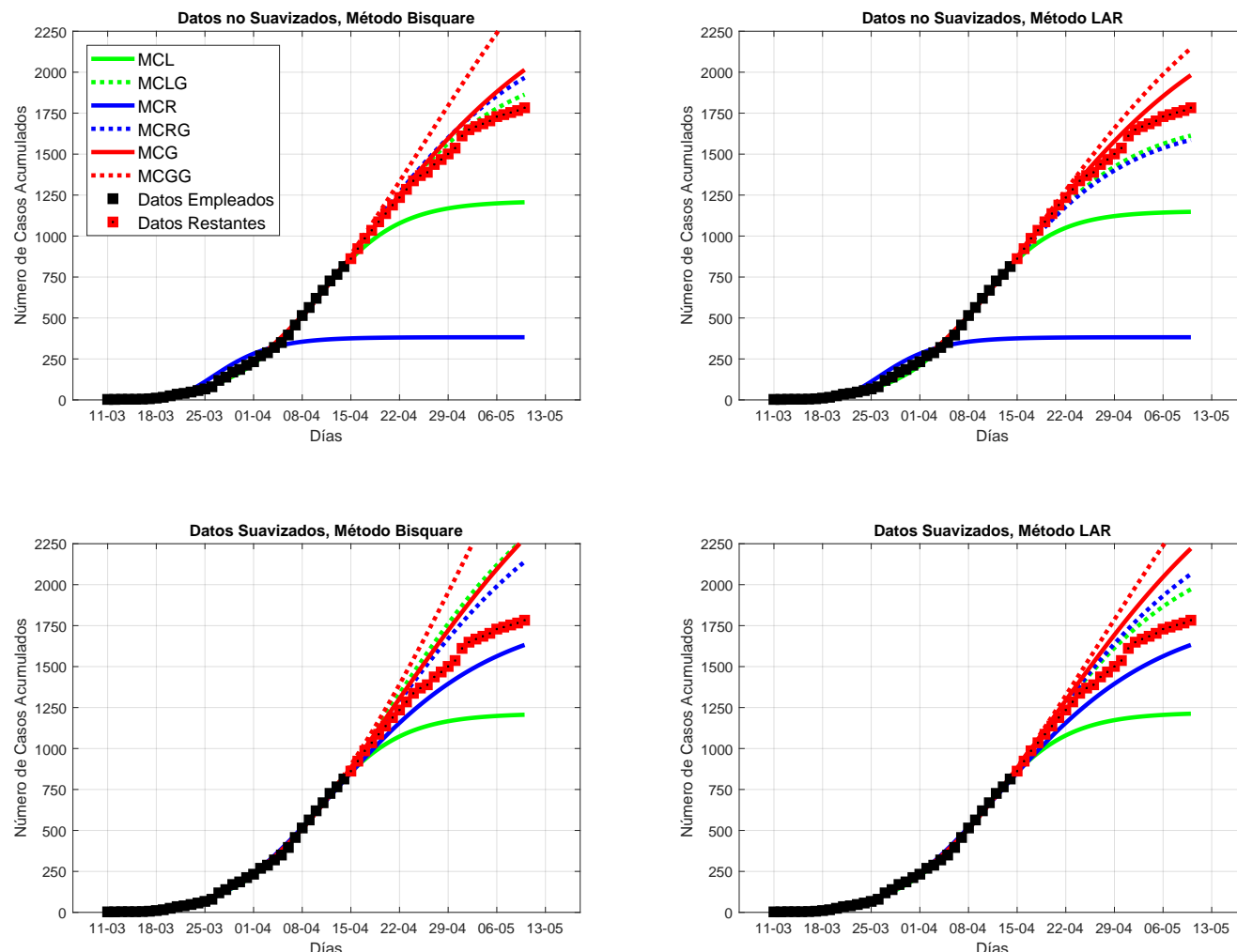


Figura 6. Resultados obtenidos empleando 5 semanas de datos (35 puntos de datos).

En este trabajo hemos demostrado que los modelos fenomenológicos empleados permiten un pronóstico acertado de la evolución de la COVID-19 en Cuba. Un aspecto novedoso de la presente contribución consiste en el empleo de datos suavizados. En los trabajos consultados no encontramos referencias a este aspecto que nos ha proporcionado resultados satisfactorios principalmente cuando se dispone de pocos datos sobre la epidemia.

Los principales resultados del estudio realizado en esta contribución son los siguientes:

1. La epidemia de COVID-19 en Cuba puede alcanzar, en esta primera ola epidémica, un máximo de 2282 casos. Esta afirmación es válida a partir de las medidas de contención implementadas. El relajamiento de estas medidas debe ser controlado para mantener este pronóstico válido.
2. El punto de inflexión en el comportamiento de la epidemia ocurrió entre los días comprendidos del 14 al

17 de Abril. El punto de inflexión, que a diferencia del llamado “pico” de la epidemia presentado por los medios de comunicación en Cuba, si posee una definición matemática bien establecida, indica que poco más de un mes después del inicio de la epidemia se abandonó la tendencia de crecimiento exponencial en el contagio.

3. Los modelos demuestran que las medidas implementadas por el gobierno y MINSAP cubanos fueron efectivas para controlar la expansión de la COVID-19. En las tablas 1 y 2 se observa claramente esta tendencia dados los valores de γ y $\bar{\mu}$ de los modelos MCE y MCEG respectivamente.
4. Emplear datos suavizados mediante media móvil en lugar de los datos originales es muy apropiado cuando se dispone de pocos datos para realizar pronósticos de comportamiento de la epidemia mediante modelos fenomenológicos. Esta afirmación puede ser válida para otros tipos de modelos como los de compartimentos y

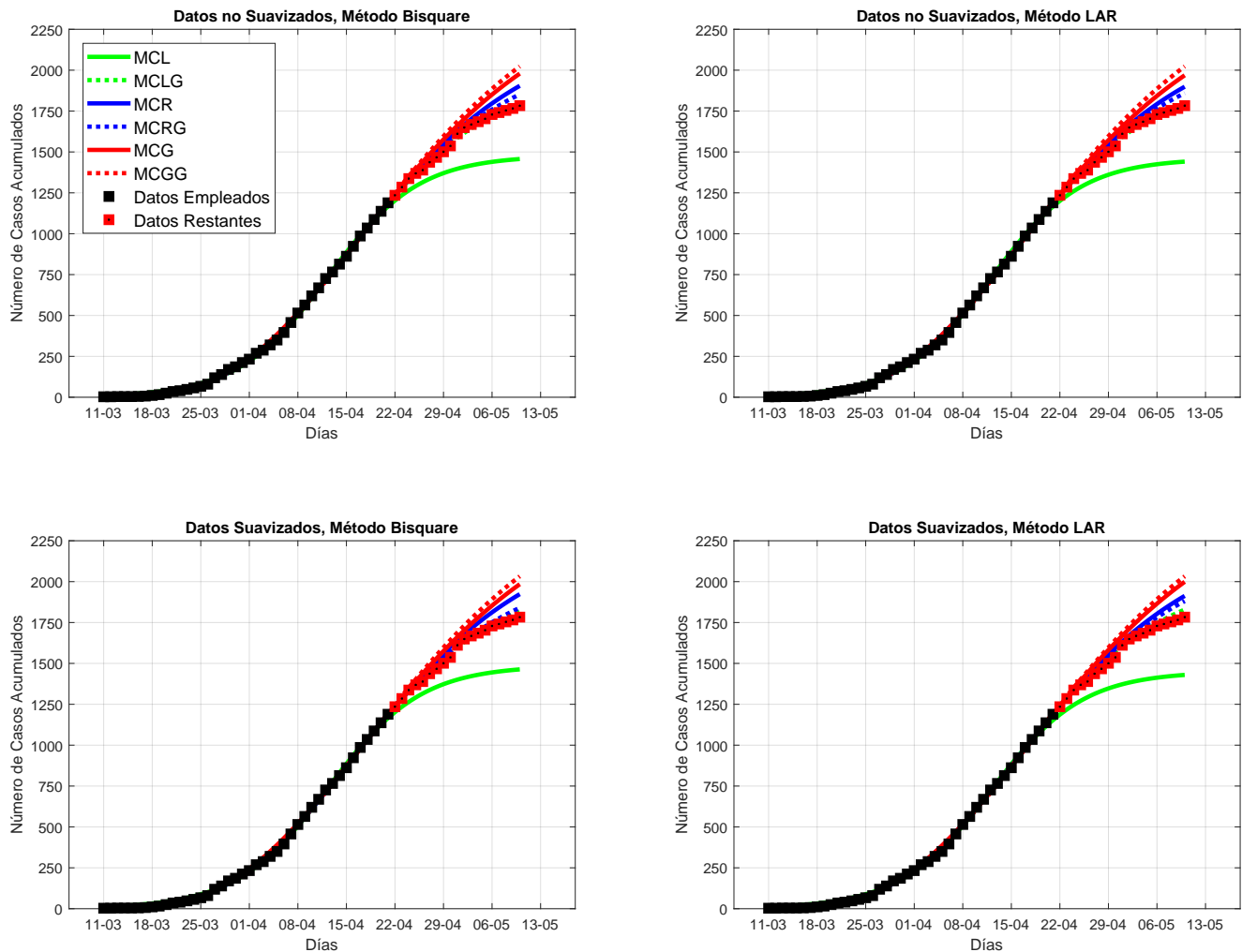


Figura 7. Resultados obtenidos empleando 6 semanas de datos (42 puntos de datos).

recomendamos su verificación. En el presente trabajo se empleó consistentemente una media móvil a 7 días, pero este valor puede ser objeto de estudios futuros.

5. En el ajuste de datos se deben emplear métodos robustos. Dado el empleo de la función fit de la Curve Fitting Toolbox de Matlab, solo se probaron los métodos denominados Bisquare y LAR. Existen otros métodos que funciones como la `nlinfit` de la Statistics and Machine Learning Toolbox de Matlab pone a disposición de los usuarios. Una experimentación más amplia en este sentido es necesaria.
6. La afirmación planteada por los autores en [30] referente a que el modelo MCL subestima el curso real de la epidemia y de que el MCG por el contrario la sobrestima, es cierta para la COVID-19 en Cuba cuando el conjunto de datos tiene más de 28 valores. Con menos valores no se cumple esta afirmación. Un trabajo futuro interesante, que también pretendemos abordar, es

comprobar esta conjetura con datos de la COVID-19 en todo el planeta. En general pretendemos aplicar las experiencias alcanzadas con este estudio en Cuba a todo el planeta.

En trabajos futuros nos proponemos abordar las interrogantes planteadas en los puntos 4, 5 y 6 anteriores, así como aplicar métodos de suavizado de datos en un modelo simple de compartimentos como el SIR o su variante SIOR. También nos proponemos explorar cuestiones relevantes que no hemos abordado en el presente trabajo como la sensibilidad de los modelos a los parámetros y otras vías de cuantificar la incertidumbre de las predicciones resultantes (vea [8]).

Referencias

- [1] Alberti, Tommaso y Davide Faranda: *On the uncertainty of real-time predictions of epidemic growths: a COVID-19 case study for China and Italy*. arXiv preprint arXiv:2004.10060, páginas 1–24, 2020.

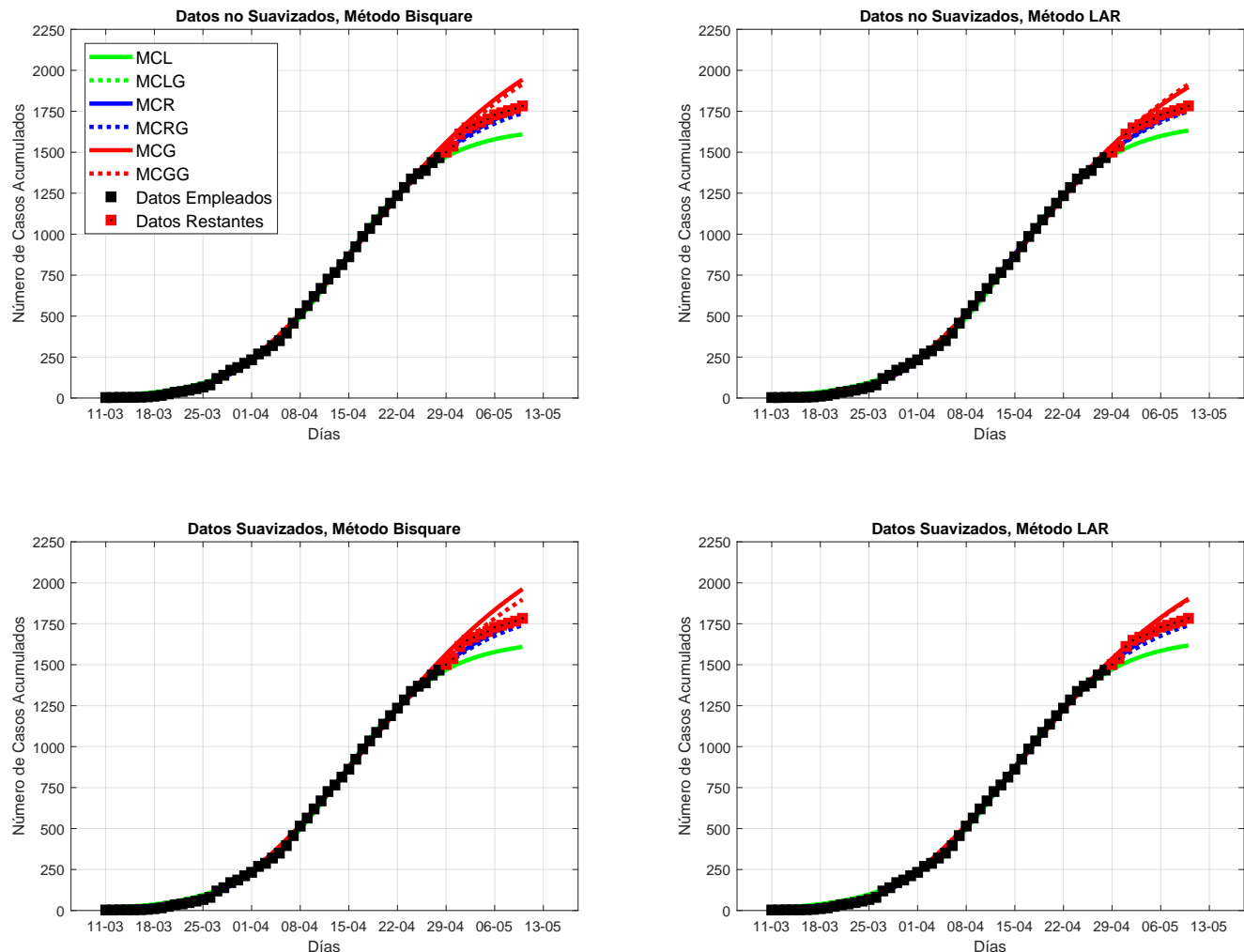


Figura 8. Resultados obtenidos empleando 7 semanas de datos (49 puntos de datos).

- [2] Bürger, Raimund, Gerardo Chowell y Leidy Y. Lara-Díaz: *Comparative analysis of phenomenological growth models applied to epidemic outbreaks*. Mathematical Biosciences and Engineering, 16(5):4250–4273, 2019.
- [3] Casella, Francesco: *Can the COVID-19 epidemic be managed on the basis of daily data?* arXiv:2003.06967 [physics, q-bio], 2020.
- [4] Castorina, P., A. Iorio y D. Lanteri: *Data analysis on Coronavirus spreading by macroscopic growth laws*. arXiv preprint arXiv:2003.00507, 2020.
- [5] Chen, Dongmei, Bernard Moulin y Jianhong Wu: *Analyzing and Modeling Spatial and Temporal Dynamics of Infectious Diseases*. Wiley, 2015.
- [6] Chen, Xinguang y Bin Yu: *First two months of the 2019 Coronavirus Disease (COVID-19) epidemic in China: real-time surveillance and evaluation with a second derivative model*. Global Health Research and Policy, 5(1):7, 2020.
- [7] Cherniha, Roman y Vasyl' Davydovych: *A mathematical model for the coronavirus COVID-19 outbreak*. arXiv preprint arXiv:2004.01487, páginas 1–9, 2020.
- [8] Chowell, Gerardo: *Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts*. Infectious Disease Modelling, 2:379–398, 2017.
- [9] Chowell, Gerardo, Lisa Sattenspiel, Shweta Bansal y Cécile Viboud: *Mathematical models to characterize early epidemic growth: A Review*. Physics Life Reviews, 18:66–97, 2016.
- [10] Dehkordi, Amirhoshang Hoseinpour, Majid Alizadeh, Pegah Derakhshan, Peyman Babazadeh y Arash Jahan-dideh: *Understanding Epidemic Data and Statistics: A*

- case study of COVID-19. arXiv:2003.06933 [q-bio], 2020. arXiv: 2003.06933.
- [11] Diekmann, Odo, Hans Heesterbeek y Tom Britton: *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press, 2013.
- [12] Foppa, Ivo M.: *A Historical Introduction to Mathematical Modeling of Infectious Diseases. Seminal Papers in Epidemiology*. Academic Press, 2017.
- [13] Hsieh, Ying Hen: *Richards Model: A Simple Procedure for Real-time Prediction of Outbreak Severity*. En *Modelling and Dynamics of Infectious Diseases*, páginas 216–236. World Scientific, 2009.
- [14] 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, 3:283:1–8, 2010.
- [15] Li, Michael Y.: *An Introduction to Mathematical Modeling of Infectious Diseases*. Springer, 2010.
- [16] Ma, Stefan y Yingcun Xia: *Mathematical Understanding of Infectious Disease Dynamics*. World Scientific, 2009.
- [17] Maier, Benjamin F. y Dirk Brockmann: *Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 outbreak in Mainland China*. arXiv:2002.07572 [physics, q-bio], 2020.
- [18] Richards, F.J.: *A Flexible Growth Function for Empirical Use*. Journal of Experimental Botany, 10(29):290–300, 1959.
- [19] Roosa, Kimberly, Yiseul Lee, Ruiyan Luo, Alexander Kirpich, Richard Rothenberg, James M. Hyman, Ping Yan y Gerardo Chowell: *Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, China: February 13–23, 2020*. Journal of Clinical Medicine, 9:596:1–9, 2020.
- [20] 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*. Epidemiol. Infect., 145:2313–2323, 2017.
- [21] Seidl, I. y C.A. Tisdell: *Carrying capacity reconsidered: From Malthus' population theory to cultural carrying capacity*. Ecological Economics, 31(3):395–408, 1999.
- [22] Sigl, Juliane: *Nonlinear residual minimization by iteratively reweighted least squares*. Computational Optimization and Applications, 64(3):755–792, 2016.
- [23] Singer, H.M.: *The COVID-19 pandemic: growth patterns, power law scaling, and saturation*. arXiv preprint arXiv:2004.03859, páginas 1–10, 2020.
- [24] 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*. Public Library of Science ONE (PLOS ONE), 12(6):1–17, 2017.
- [25] Verhulst, P.F.: *A Note on the Law of Population Growth*. En *Mathematical Demography*, páginas 333–337. Springer-Verlag, 1977.
- [26] Viboud, Cécile, Lone Simonsen y Gerardo Chowell: *A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks*. Epidemics, 15:27–37, 2016.
- [27] Vynnycky, Emilia y Richard White: *An Introduction to Infectious Disease Modelling*. Oxford University Press, 2010.
- [28] 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*. preprint, Infectious Diseases (except HIV/AIDS), 2020.
- [29] Yan, Ping y Gerardo Chowell: *Quantitative Methods for Investigating Infectious Disease Outbreaks*. Springer, 2019.
- [30] Yang, Wuyue, Dongyan Zhang, Liangrong Peng, Changjing Zhuge, Liu Hong y Zhou Pei-Yuan: *Rational evaluation of various epidemic models based on the COVID-19 data of China*. arXiv preprint arXiv:2003.05666, páginas 1–18, 2020.
- [31] Zhao, Shi, Salihu S. Musa, Hao Fu, Daihai He y Jing Qin: *Simple framework for real-time forecast in a data-limited situation: the Zika virus (ZIKV) outbreaks in Brazil from 2015 to 2016 as an example*. Parasites & Vectors, 12:1–13, 2019.