

# Predicciones en tiempo real de la Covid-19 en Cuba y sus provincias: Modelos fenomenológicos dinámicos y un modelo no lineal con efectos mixtos. Real-time predictions of Covid-19 in Cuba and its provinces: Dynamic phenomenological models and a nonlinear mixed effects model

Carlos R. Sebrango Rodríguez<sup>1\*</sup>, Lizet Sánchez Valdés<sup>2</sup>, Waldemart Baldoquin<sup>3</sup>, Osvaldo Norman Montenegro<sup>4</sup>, Raúl Guinovart<sup>5</sup>

**Resumen** En ausencia de información confiable sobre los mecanismos de transmisión de una infección emergente como la Covid-19, modelos fenomenológicos dinámicos pueden proporcionar una evaluación temprana de un posible alcance del brote en tiempo real. En este estudio se emplean diferentes modelos fenomenológicos dinámicos, que tienen en cuenta diferentes perfiles de crecimiento inicial de la epidemia: los modelos logístico y de Richards generalizados, para generar pronósticos de un horizonte de 10 días de la incidencia de casos confirmados de la Covid-19 en Cuba, haciendo uso de un enfoque de bootstrap paramétrico para cuantificar la incertidumbre. El pronóstico del número de casos confirmados a nivel provincial, se realiza usualmente para cada provincia por separado, ignorando la variabilidad entre y dentro de las provincias. En la presente investigación se hacen pronósticos con un horizonte de 7 días de los casos confirmados de cada una de las provincias de Cuba utilizando un modelo no lineal con efectos mixtos que tiene en cuenta la heterogeneidad entre las provincias.

**Abstract** In the absence of reliable information about transmission mechanisms of an emerging infection, simple phenomenological models can provide an early assessment of the potential scope of outbreaks in near real-time. In this study, several dynamic phenomenological models are used, that incorporate flexible early epidemic growth profiles, the generalized logistic and Richards models, in order to generate 10-day ahead forecasts of the incidence of confirmed cases of Covid-19 in Cuba. In addition, we use a parametric bootstrap approach to quantify the uncertainty. Short-term forecasts at the provincial level are usually performed for each province separately, ignoring variability between and within provinces. In this research, we generate 10-day ahead forecasts of cumulative case reports of each province using a nonlinear mixed effects model that takes into account the heterogeneity between provinces.

## Palabras Clave

Covid-19, modelos fenomenológicos dinámicos, modelo no lineal con efectos mixtos

## Keywords

Covid-19, dynamic phenomenological model, Nonlinear mixed effects model

<sup>1</sup> Universidad de Sancti Spiritus "José Martí Pérez", Centro de Estudios de Energía y Procesos Industriales (CEEPI), Cuba, sebrango@uniss.edu.cu

<sup>2</sup> Centro de Inmunología Molecular, La Habana, Cuba, lsanchez@cim.sld.cu

<sup>3</sup> Instituto de Medicina Tropical IPK, La Habana, Cuba, wbaldoquin@ipk.sld.cu

<sup>4</sup> Centro de Bioactivos Químicos, UCLV, Villa Clara, Cuba, onorman@uclv.edu.cu

<sup>5</sup> Universidad de La Habana, Facultad Matemática y computación, Cuba, gino@matcom.uh.cu

\* Autor para Correspondencia, Corresponding Author

## Introducción

La actual epidemia de una nueva enfermedad por coronavirus (Covid-19) comenzó en la provincia de Hubei, China, en

diciembre de 2019 y se ha expandido rápidamente llegando a 185 países. Cuba tampoco ha podido escapar de este mortal virus, el 11 de marzo del 2020 tuvo sus primeros casos

confirmados. En el país se han establecido políticas públicas efectivas, para contener la epidemia Covid-19: aislamiento, cuarentenas, distanciamiento social, reordenamiento laboral, restricciones de viajes, etc. A pesar de estas políticas el número de casos sigue aumentando y el país necesita tomar decisiones sobre diferentes recursos para lograr controlar la epidemia. La modelación matemática puede jugar un papel fundamental en este fin.

En los últimos años ha existido un interés creciente en el uso de los modelos matemáticos para el análisis de brotes de enfermedades infecciosas [4]. Estos modelos han permitido describir el comportamiento de brotes, facilitar la estimación de parámetros epidemiológicos claves, evaluar el impacto de las intervenciones de control, hacer predicciones en tiempo-real y ayudar a entender y describir la dinámica de la transmisión. Cuando la comprensión del impacto de una enfermedad infecciosa emergente, como la Covid-19, es limitada, el uso de modelos fenomenológicos simples puede proporcionar una evaluación temprana de un posible alcance del brote en tiempo casi real.

El desarrollo de herramientas estadísticas para la estimación y predicción de parámetros en epidemias de enfermedades infecciosas emergentes y reemergentes representa una prioridad para las autoridades de salud en su trabajo para prevenir y mitigar las amenazas de enfermedades [18]. Los modelos fenomenológicos pueden ayudar mucho en este sentido. Los modelos fenomenológicos (por ejemplo, Hsieh y cols. [10]) prescindir de los supuestos de los modelos matemáticos de transmisión. En su lugar, ellos hacen supuestos generales acerca de la forma de la curva de incidencia. En general, tales modelos fenomenológicos utilizan pocos parámetros, hacen suposiciones más sencillas, y son probables a dar estimaciones más robustas cuando son aplicados en contextos con datos muy limitados [13].

Entre los modelos fenomenológicos más utilizado para la modelación de datos de enfermedades infecciosas se encuentra el modelo logístico [7, 19, 6] y el modelo de Richards [14, 20, 11, 9, 10, 8], aunque recientemente Chowell y cols. [3, 2, 5, 4, 1] han utilizado una generalización de estos modelos, el modelo logístico y de Richards generalizados, que ha ayudado a caracterizar los perfiles de crecimiento epidémico temprano a partir de datos de brotes de enfermedades infecciosas como Zika, Ébola y VIH SIDA. Estos trabajos previos han demostrado que los modelos fenomenológicos dinámicos, puede capturar los patrones empíricos de epidemias pasadas y son útiles para generar pronósticos a corto plazo de la trayectoria epidémica en tiempo real. Los pronósticos a corto plazo en tiempo real generados a partir de tales modelos pueden ser útiles para guiar la asignación de recursos necesarios para mitigar o controlar la epidemia.

Algunas veces los reportes de casos acumulados se coleccionan para todas las áreas urbanas en una región en particular, y para los epidemiólogos resulta de interés estimar los parámetros epidemiológicos claves o estudiar la heterogeneidad regional de la transmisión de la enfermedad. La modelación,

para estimar parámetros epidemiológicos y hacer pronósticos a corto plazo, se realiza usualmente para cada área por separado, pero cuando el interés consiste en estimar el comportamiento promedio de un área en particular en la población, y la variabilidad entre y dentro de las áreas, se propone un modelo no lineal con efectos mixtos [17].

En este estudio se emplean diferentes modelos fenomenológicos para generar pronósticos de un horizonte de 10 días de la incidencia y del número acumulado de casos confirmados de la Covid-19 en Cuba, a nivel nacional y provincial respectivamente. Para el pronóstico de la incidencia casos confirmados en Cuba y para estudiar los perfiles de crecimiento epidémico temprano, se utilizan los modelos logístico y de Richards generalizados. Para el pronóstico a corto plazo de los casos confirmados en las provincias, se utiliza un modelo no lineal con efectos mixtos que tiene en cuenta la variabilidad entre todas las provincias.

El artículo está organizado de la siguiente forma: La Sección 1 Materiales y métodos, describe los datos utilizados, se presentan los modelos fenomenológicos dinámicos, la cuantificación de la incertidumbre y el modelo no lineal con efectos mixtos para la modelación y pronóstico. Los resultados de los métodos propuestos a los datos se muestran en la Sección 2 y en la Sección 3 se discuten los resultados y se hacen las conclusiones.

## 1. Materiales y Métodos

### 1.1 Fuente de datos

En el presente estudio se utilizaron los datos de casos confirmados diarios de Covid-19 desde el 11 de marzo del 2020 hasta el 21 de mayo del 2020 para toda Cuba y las 15 provincias y el municipio especial Isla de la Juventud. Esta serie de datos fueron obtenidos a partir de los reportes oficiales emitidos por parte de las autoridades del ministerio de salud pública y del sitio Web <https://covid19cubadata.github.io>.

### 1.2 Modelos fenomenológicos para caracterizar perfiles de crecimiento epidémico temprano

En las epidemias emergentes, los datos epidemiológicos son limitados, y los parámetros epidemiológicos necesarios para calibrar modelos elaborados de transmisión mecanicista aún no se han aclarado completamente. Los pronósticos a corto plazo en tiempo real deben basarse en modelos fenomenológicos dinámicos que hayan sido validados durante brotes de enfermedades infecciosas previas [3, 2, 5, 4, 1, 15, 16]

#### 1.2.1 Modelo logístico generalizado

El modelo logístico generalizado (MLG) es una extensión del modelo logístico que incluye un parámetro adicional,  $p$ , para permitir la ampliación del crecimiento;  $p = 1$  indica un crecimiento exponencial temprano,  $p = 0$  representa un crecimiento constante, y  $0 \leq p \leq 1$  acomoda crecimientos subexponenciales o polinómicos tempranos [3, 2, 5, 4, 1, 15, 16]. El GLM está definido por la ecuación diferencial:

$$C'(t) = rC^p(t) \left[ 1 - \frac{C(t)}{K} \right]$$

donde  $C(t)$  representa el número de casos acumulados en el tiempo  $t$ ,  $C'(t)$  representa el número de nuevos casos infectados en el tiempo  $t$ ,  $r$  es la razón de crecimiento,  $p$  es la escala del parámetro de crecimiento y  $K$  es la capacidad de carga o el tamaño final de la epidemia.

### 1.2.2 Modelo de Richards generalizado

El modelo de Richards generalizado (MRG) es una extensión del modelo de Richards, que asume una fase de crecimiento inicial exponencial. Para tener en cuenta dinámicas de crecimiento inicial [3, 2, 5, 4, 1, 15, 16], se creó el MRG que incorpora el parámetro  $p$ , también conocido parámetro de desaceleración del crecimiento. Por lo que, el MRG tiene la forma:

$$C'(t) = rC^p(t) \left[ 1 - \left( \frac{C(t)}{\alpha} \right)^a \right]$$

Donde  $0 \leq p \leq 1$ . En las etapas iniciales de la epidemia, este modelo es capaz de captar diferentes perfiles de crecimientos, mencionados con el modelo anterior.

Las estimación y predicciones de los parámetros de estos modelos se generaron empleando la función LSQCURVEFIT del MATLAB (The Mathworks, Inc.) con la opción Levenberg-Marquardt. Para cuantificar la incertidumbre en la estimación de los parámetros del modelo y en el pronóstico del número de casos se construyeron intervalos de confianza como en estudios anteriores [3, 2, 5, 4, 1, 15, 16] mediante la simulación de 200 realizaciones de la curva de mejor ajuste utilizando un bootstrap paramétrico con una estructura de error de Poisson. Los intervalos de confianza del 95 % se calcularon tomando los percentiles 2.5 y 97.5 de las distribuciones de parámetros generados.

Un modelo propiamente calibrado a los datos puede usarse para generar predicciones a corto plazo o a largo plazo del sistema, generando una predicción basada en la incertidumbre de modelos dada por  $f(t, \hat{\theta}_1), f(t, \hat{\theta}_2), \dots, f(t, \hat{\theta}_{200})$  (ajustes de las 200 curvas generadas por el procedimiento de cuantificación de la incertidumbre), la cual requiere propagar la incertidumbre del estado actual a un horizonte de tiempo de  $h$  unidades:  $f(t+h, \hat{\theta}_1), f(t+h, \hat{\theta}_2), \dots, f(t+h, \hat{\theta}_{200})$ .

Si bien podemos inspeccionar los residuos para detectar cualquier desviación sistemática del modelo que se ajuste a los datos, también es posible cuantificar el error del modelo ajustado utilizando métricas de rendimiento. Estas métricas son útiles también para cuantificar el error asociado en los pronósticos. Entre algunas de las métricas más utilizadas están la suma de cuadrados del error (SSE), error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE) [1].

### 1.3 Modelo fenomenológico con efectos mixtos

Algunas veces los reportes de casos acumulados se coleccionan para todas las provincias de un país en particular, y para los epidemiólogos resulta de interés estimar, pronosticar parámetros epidemiológicos claves o estudiar la heterogeneidad

regional de la transmisión de la enfermedad. La modelación, se realiza usualmente para cada provincia por separado, pero cuando el interés consiste en tener en cuenta la variabilidad entre y dentro de las provincias, se propone un modelo no lineal con efectos mixtos. En esta sección se describe el modelo fenomenológico con efectos mixtos

El modelo no lineal con efectos mixtos correspondiente para el número acumulado de casos reportados  $Y_{ij}$  en la provincia  $i$  en el tiempo  $t_j$  es

$$Y_{ij} = \frac{\alpha_i}{\left[ 1 + k_i \cdot e^{-k_i \cdot \gamma_i (t_j - \eta_i)} \right]^{\frac{1}{k_i}}} + \varepsilon_{ij}, \quad (1)$$

El vector de parámetros para la provincia específica es modelado como:

$$\theta_i = X_i \theta + Z_i b_i \quad (2)$$

Donde  $\theta$  es el vector de parámetros fijos,  $b_i$  es el vector de efectos aleatorios de la provincia específica,  $X_i$  y  $Z_i$  son matrices de diseño conocido para los efectos fijos  $\theta$  y los efectos aleatorios  $b_i$ , respectivamente:

$$X_i = Z_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \theta = \begin{pmatrix} \alpha \\ k \\ \gamma \\ \eta \end{pmatrix}, b_i = \begin{pmatrix} b_{i1} \\ b_{i2} \\ b_{i3} \\ b_{i4} \end{pmatrix} \quad (3)$$

De la ecuación 2 sigue que el vector de parámetros de área específica puede ser expresada como

$$\begin{pmatrix} \alpha_i \\ k_i \\ \gamma_i \\ \eta_i \end{pmatrix} = \begin{pmatrix} \alpha + b_{i1} \\ k + b_{i2} \\ \gamma + b_{i3} \\ \eta + b_{i4} \end{pmatrix} \quad (4)$$

Se asume que los efectos aleatorios  $b_i$  están distribuidos normalmente como  $b_i \sim N(0, \Psi)$  y el error dentro del grupo  $\varepsilon_{ij} \sim N(0, \sigma^2)$ . Para representar la estructura de varianza-covarianza de los efectos aleatorios  $\Psi$  se utiliza una matriz general definida positiva.

Para la estimación de los parámetros y pronóstico del número de casos confirmados en cada una de las provincias por el modelo no lineal con efectos mixtos se utiliza el paquete **nlme** del software R. La función **nlme** de ese paquete implementa el algoritmo propuesto por Lindstrom y Bates [12]. Este algoritmo de estimación alterna entre dos pasos, un paso de mínimos cuadrados no lineales penalizado y un paso de efectos mixtos lineales.

#### 1.3.1 Estrategia de construcción del modelo con efectos mixtos

En esta sección, se explica brevemente la estrategia que se debe seguir para obtener el modelo no lineal con efectos mixtos que mejor ayude a explicar la variabilidad de los parámetros epidemiológicos dentro de las provincias.

Un paso crucial en la construcción de modelos con efectos mixtos es decidir que coeficientes en el modelo necesitan

efectos aleatorios para tener en cuenta la variación entre las provincias y cuales deben ser tratados como efectos puramente fijos. Para ello, la estrategia de modelación es ajustar varios modelos teniendo en cuenta en cada modelo diferentes parámetros con efectos aleatorios. Dentro de la estrategia, una alternativa es considerar en primer lugar un modelo sin efectos aleatorios. Luego, se deben utilizar todas las combinaciones de parámetros con efectos aleatorios. En esta alternativa se sugiere asumir que la matriz de varianza-covarianza tiene una estructura general definida positiva.

Otra estrategia alternativa de construcción del modelo es comenzar con un modelo con efectos aleatorios para todos los parámetros y entonces examinar los resultados del ajuste para decidir cuál de los efectos aleatorios debe ser eliminado del modelo. Un problema con este enfoque es que, cuando se asume una estructura general definida positiva para la matriz de varianza-covarianza de los efectos aleatorios, el número de parámetros a estimar crece con el cuadrado del número de efectos aleatorios. En casos donde el número de efectos aleatorios es relativamente grande al número de grupos (provincias), se recomienda generalmente utilizar inicialmente una estructura de matriz diagonal para la matriz de varianza-covarianza de los efectos aleatorios, para prevenir problemas de convergencia con un modelo sobre-parametrizado.

Durante este proceso, se sugiere observar en las salidas de los resultados, las estimaciones de las desviaciones estándar de los efectos aleatorios ya que desviaciones cerca de cero pueden sugerir que ese parámetro debe ser considerado como efecto fijo. Uno puede probar si cierto efecto aleatorio puede ser eliminado del modelo, realizando la prueba de razón de verosimilitud.

Otro aspecto importante a observar en los resultados, es la matriz de correlaciones estimadas de los efectos aleatorios porque puede sugerir la estructura de la matriz de varianza-covarianza. Por ejemplo, altas correlaciones entre dos efectos aleatorios comparados con otras correlaciones de otros efectos aleatorios sugiere que la matriz es no condicionada y que la estructura de los efectos aleatorios puede estar sobre-parametrizada, quizás sugiriendo una estructura diagonal en bloque. Un gráfico de diagnóstico útil para evaluar la sobre-parametrización, es el gráfico de dispersión de los efectos aleatorios estimados que aparecen en muchos software, en R se puede utilizar la función **pairs** del paquete **nlme**. También se puede utilizar la prueba de razón de verosimilitud para probar si cierta estructura de matriz de varianza-covarianza debe ser considerada.

Finalmente, se debe chequear si los supuestos del modelo no lineal con efectos aleatorio parecen válidos para el modelo ajustado a los datos. Para ellos el uso de métodos gráficos proporcionan las herramientas más útiles para evaluar esos supuestos. El gráfico de los errores estandarizados contra los valores ajustados, ayuda a evaluar el supuesto que los errores dentro del grupo son independientes e idénticamente distribuidos con distribución normal con media cero y varianza  $\sigma^2$ . Si los residuos están distribuidos simétricamente alrededor de cero con una varianza aproximadamente constante, esto indi-

caría que no hay ninguna violación de este supuesto. Si hay un patrón diferente, no se cumpliría este supuesto y sería necesario, quizás, utilizar funciones de la varianza (implementadas en el paquete **nlme** de R) para modelar la heterocedasticidad. El gráfico de los errores estandarizados contra los cuantiles de la distribución normal ayuda a evaluar la violación del supuesto de normalidad para los errores dentro de los grupos. La adecuación del modelo ajustado se debe visualizar mejor realizando un gráfico conjunto con los valores observados y ajustados o predichos, donde se observen las predicciones entre las provincias (usando los efectos aleatorios estimados). Un aspecto a tener en cuenta, es que el modelo final obtenido debe tener sentido práctico.

## 2. Resultados

### 2.1 Modelos fenomenológicos para caracterizar perfiles de crecimiento epidémico temprano

Para estudiar los perfiles de crecimiento epidémico y hacer pronósticos a corto plazo se ajustaron el modelo logístico generalizado (MLG) y el modelo de Richards generalizado (MRG) a los datos de los primeros 25, 30, 35, 40, 45 y 50 días de la epidemia Covid-19 en Cuba. En las tablas 1 y 2 se muestran diferentes estadísticos o métricas de rendimiento MAE, RMSE y SSE del ajuste de estos dos modelos para los diferentes intervalos de datos respectivamente. Observe como las métricas del modelo de Richards generalizado son ligeramente superiores a los del modelo logístico generalizado.

Las estimaciones de los parámetros  $p$  y  $K$  para los dos modelos fenomenológicos dinámicos estudiados, con incertidumbre cuantificada generada por la metodología descrita en la sección 1.2 se muestra en las figuras 1 y 2 respectivamente. En estas figuras se muestran los histogramas de las distribuciones empíricas de las estimaciones de parámetros  $p$  y  $K$  utilizando 200 realizaciones bootstrap para cada uno de los períodos de datos epidémicos estudiados. Además, para cada período en el panel de abajo se muestran las 200 realizaciones de la curva epidémica asumiendo una estructura de error de Poisson y las líneas de mejor ajuste y los intervalos de confianzas del 95% de estos dos modelos MLG y MRG en su período de calibración.

Es importante destacar que tanto en las figuras 1 y 2 con solo 25 días de datos, no es posible estimar de manera confiable el parámetro de desaceleración de crecimiento  $p$ , porque su intervalo de confianza varía de 0.7 a 1. De hecho, no es posible discriminar entre dinámicas subexponenciales y de crecimiento exponencial basados en solo los primeros 25 datos. A medidas que se emplean más datos de la fase de crecimiento inicial para estimar los parámetros de los modelos, la incertidumbre en las estimaciones de los parámetros no solo se reduce sino que las estimaciones de los parámetros están más restringidas alrededor de sus valores verdaderos. Sin embargo, es de destacar que para el modelo de Richards generalizado con el incremento de los datos, no se comportó de esa forma ya que los intervalos de confianza siguen variando entre 0.7 y 1.

**Tabla 1.** Estadísticos o métricas de rendimiento (MAE, RMSE y SSE) del ajuste obtenido por el modelo logístico generalizado utilizando diferentes intervalos de datos epidémicos.

Período	1-25	1-30	1-35	1-40	1-45	1-50
MAE	3.76	4.45	4.53	4.87	4.77	5.16
RMSE	5.74	6.86	7.07	7.08	6.79	7.35
SSE	824.92	1412.58	1747.19	2007.02	2073.11	2703.15

**Tabla 2.** Estadísticos o métricas de rendimiento (MAE, RMSE y SSE) del ajuste obtenido por el modelo de Richards generalizado utilizando diferentes intervalos de datos epidémicos.

Periodo	1-25	1-30	1-35	1-40	1-45	1-50
MAE	3.75	4.28	4.53	4.87	4.66	5.19
RMSE	5.73	6.81	6.71	7.08	6.77	7.35
SSE	820.71	1389.44	1576.08	2007.02	2059.63	2701.11

Para las estimaciones del parámetro  $K$ , tamaño final de la epidemia, ocurre algo similar. Con solo pocos datos, sobre todos para períodos de datos que no incluyen el pico hay mucha incertidumbre. Sin embargo, para los períodos con 40, 45 y 50 datos ambos modelos, MLG y MRG, estiman este parámetro de manera más confiable, sus estimaciones están alrededor de los 2000 casos confirmados.

Estos modelos fenomenológicos dinámicos pueden ser utilizados para el pronóstico a corto plazo, más específicamente para el pronóstico de un horizonte de 10 días de la incidencia de casos confirmados. Las figuras 3 y 4 muestran los pronósticos de 10 días basados en el modelo logístico y modelo de Richards generalizados respectivamente. Estos modelos fueron calibrados utilizando una cantidad creciente de datos epidémicos (25, 30, 35, 40, 45 y 50 días). Observe como para intervalos de tiempos antes del pico los pronósticos subestiman la incidencia de casos o la incertidumbre de los pronósticos es considerable. Sin embargo, a medida que los períodos de datos para la calibración sobrepasan el pico de la epidemia, los pronósticos se hacen mas confiables, la banda de confianza se hace más estrecha considerando ambos modelos.

## 2.2 Modelo fenomenológico con efectos mixtos

Nuestra estrategia de modelación fue ajustar diferentes modelos teniendo en cuenta en cada modelo diferentes parámetros con efectos aleatorios. Dentro de la estrategia, al principio no se consideraron parámetros con efectos aleatorios. Luego, se utilizaron todas las combinaciones de parámetros con efectos aleatorios y se consideró como la estructura de varianza-covarianza de los efectos aleatorios  $\Psi$  la matriz general definida positiva. En todos los casos, el parámetro  $\alpha$ , tamaño final de la epidemia, se consideró como un efecto aleatorio por su gran variabilidad. La tabla 3 muestra las estimaciones de los parámetros para el modelo no lineal con efectos aleatorios de mejor ajuste. Los menores valores de  $AIC = 5551,842$  y  $BIC = 5590,036$  fueron obtenidos para el modelo con los

parámetros  $\alpha$ ,  $k$  y  $\eta$  como efectos aleatorios.

**Tabla 3.** Estimación de los parámetros para el modelos con  $\alpha$ ,  $\eta$ ,  $\gamma$  como efectos aleatorios con matriz de varianza-covarianza bloque-diagonal.

Parámetro	Estimación (IC 95 %)
Efectos fijos:	
$\alpha$	121.59 (10.00, 233.17)
$k$	0.71 (0.63, 0.79)
$\gamma$	0.20 (0.15, 0.24)
$\eta$	32.85 (30.12, 35.57)
Varianza	
$\sigma^2$	23.65 (21.47, 26.04)
Efectos aleatorios	
$SD(b_{i1})$	227.39 (160.75, 321.63)
$SD(b_{i3})$	0.075 (0.05, 0.11)
$SD(b_{i4})$	5.17 (3.57, 7.49)

La matriz del gráfico de dispersión de los efectos aleatorios proporciona un diagnóstico útil para evaluar los problemas de sobre-parametrización. La no existencia de patrones entre los efectos aleatorios  $\alpha$ ,  $\eta$  y  $\gamma$  (Figura 5, Panel izquierdo) indica que el modelo no está sobre-parametrizado. Las bajas correlaciones entre los efectos aleatorios sugieren que la matriz diagonal  $\Psi$  pudiera ser utilizada para representar la estructura de varianza-covarianza de los efectos aleatorios. Para probar si una matriz diagonal  $\Psi$  pudiera ser utilizada para representar la estructura de varianza-covarianza de los efectos aleatorios, se llevó a cabo una prueba de razón de verosimilitud (tabla 4). El valor no significativo del  $p$ -value de la prueba de razón de verosimilitud y los valores más pequeños del  $AIC$  y  $BIC$  corroboran la estructura de varianza-covarianza diagonal.

El gráfico de los errores estandarizados contra los valores ajustados correspondiente al modelo con  $\alpha$ ,  $\eta$  y  $\gamma$  como efectos aleatorios y con matriz de varianza-covarianza  $\Psi$  diagonal presentado en el panel de la derecha de la figura 5, muestra

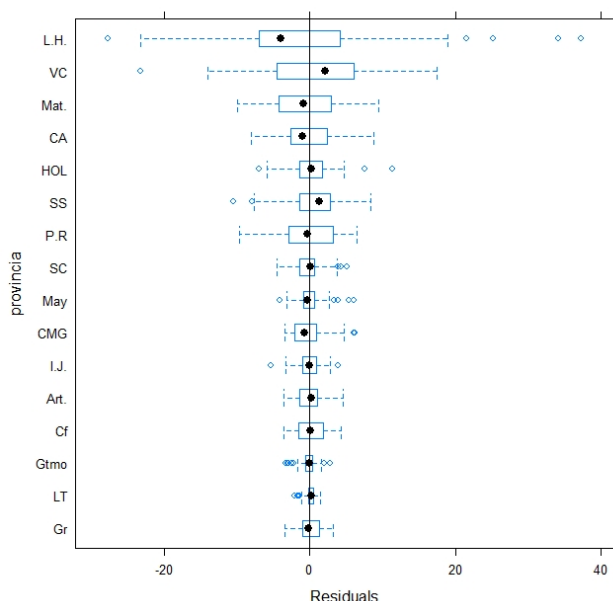


**Tabla 4.** Prueba de razón de verosimilitud para el modelo con matriz de varianza-covarianza  $\Psi$  definida positiva contra el modelo con matriz de varianza-covarianza bloque-diagonal.

Modelo	AIC	BIC	Log Ver.	Test	Razón V.	p-value
Modelo 1 ( $\Psi$ definida positiva)	5552.612	5605.128	-2765.306			
Model 2 ( $\Psi$ Bloque-diagonal)	5551.842	5590.036	-2767.921	1 vs 2	5.2303	0.1557

que los residuos están distribuidos simétricamente alrededor del cero con una varianza aproximadamente constante. Esto no indica ninguna violación de los supuestos del modelo no lineal con efectos mixtos, excepto para algunas observaciones para el provincia de La Habana.

Una de las ventajas de este tipo de modelo es que permite detectar provincias con un comportamiento atípico o que presentan mayor variabilidad de los residuos del modelo con respecto a otras. En la figura 6 se observa como las provincias más atípicas han sido La Habana, Villa Clara y Matanzas.



**Figura 6.** Análisis de los residuos del modelo para cada provincia con el fin de detectar áreas atípicas

Una evaluación final de la capacidad predictiva de este modelo está dado por el gráfico de las predicciones y pronóstico en la figura 7. Para comparación y para mostrar como los efectos individuales son tenidos en cuenta por el modelo no lineal con efectos mixtos, se presentan las predicciones dentro de las provincias (obtenidas utilizando los efectos aleatorios estimados). Note que las predicciones dentro de los grupos están en acuerdo con los casos acumulados observados, ilustrando que el modelo no lineal con efectos mixtos puede acomodar los efectos individuales. Se tomó como período de calibración los datos hasta el 14 de mayo y se tomó como período de pronóstico el período desde el 15 de mayo hasta el 21 de mayo.

### 3. Discusión y Conclusiones

Los modelos fenomenológicos son útiles en el contexto de datos epidemiológicos limitados o cuando se enfrentan epidemias desencadenadas por nuevos patógenos, como el SARS Cov-2. En el presente estudio se utilizan dos modelos fenomenológicos dinámicos, modelos logístico y de Richards generalizados, para producir conos de pronóstico de 10 días de la incidencia de casos confirmados para Cuba utilizando intervalos crecientes de datos epidémicos. Para la estimación de los parámetros de los modelos y para el pronóstico se cuantificó la incertidumbre haciendo uso de un enfoque de bootstrap paramétrico asumiendo una estructura de error de Poisson. Se mostró que la incertidumbre de las estimaciones de parámetros se reduce y se vuelve más precisa utilizando una cantidad creciente de datos de la fase de crecimiento epidémico.

En muchos estudios previos fueron utilizados modelos fenomenológicos que en su mayoría asumen una fase de crecimiento exponencial temprana como los modelos logístico, Richards, Gompertz, etc. Una de las ventajas de utilizar los modelos fenomenológicos dinámicos expuestos en este trabajo es que permiten estudiar y caracterizar diferentes perfiles de crecimiento temprano que no pueden ser tenidos en cuenta con estos modelos fenomenológicos previos. En futuras investigaciones pretendemos hacer pronósticos usando diferentes modelos fenomenológicos con diferentes perfiles de crecimiento temprano pero utilizando técnicas de promedio de modelos, es decir que tienen en cuenta todos los modelos.

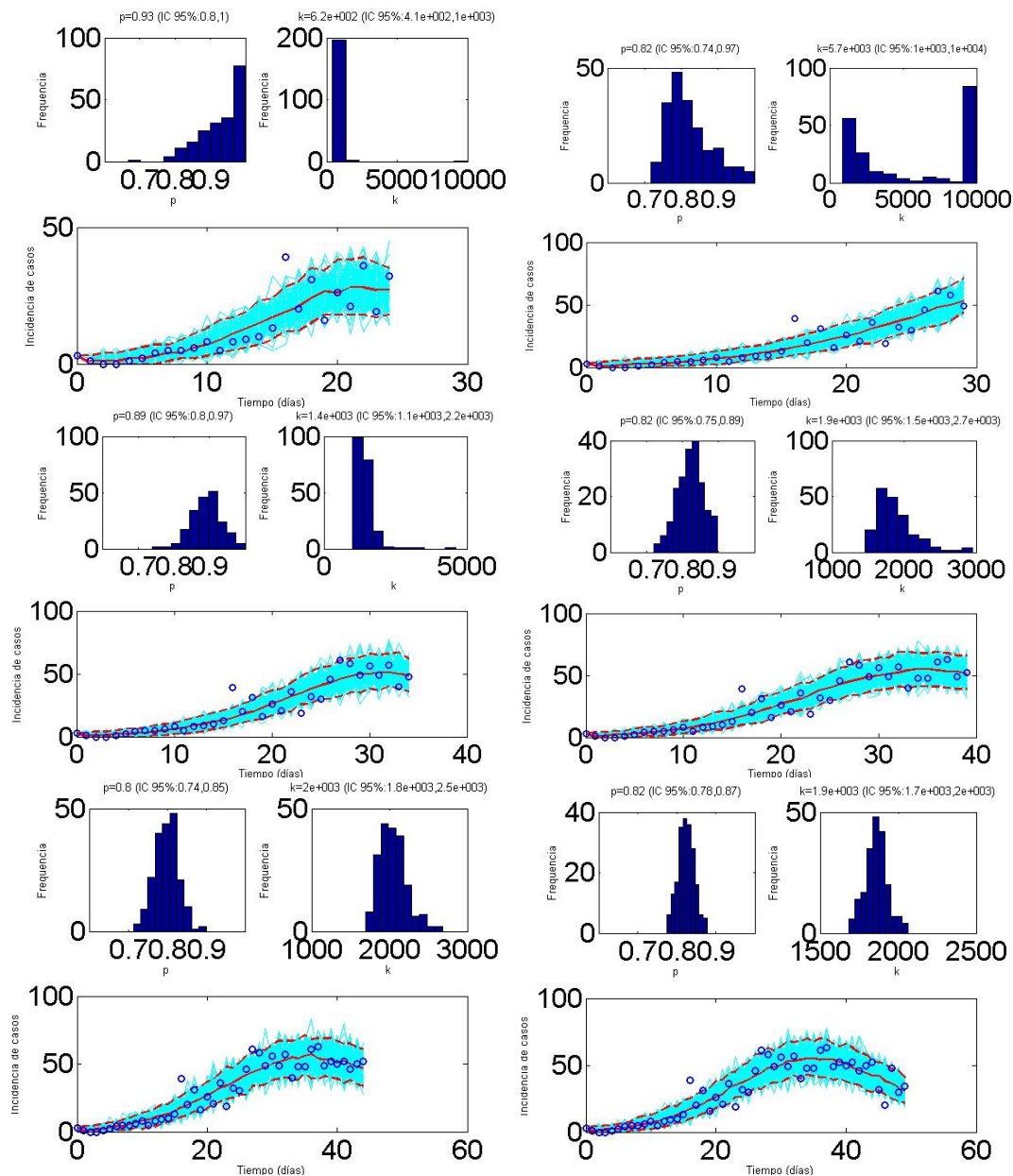
También, en la investigación se propone un modelo no lineal con efectos mixtos para modelar y hacer pronósticos de una semana para el número de casos confirmados acumulados para cada una de las provincias de Cuba. Este enfoque de modelación tiene como principal ventaja que no ignora la variabilidad entre las provincias, la cual no es tenida en cuenta cuando se modelan cada una de las provincias por separado. Además desde el punto de vista práctico, el uso de un modelo no lineal con efectos mixtos requiere de menos parámetros a estimar, permite tener en cuenta la variabilidad entre las provincias y dentro de ellas, puede ser utilizado para determinar provincias con un comportamiento atípico y permite incluir la posibilidad de incorporar variables contextuales para explicar las diferencias entre las provincias (porcentaje de pruebas PCR, densidad poblacional, variables medio ambientales, etc.). Esta última temática será objeto de estudio en futuras investigaciones.

Los pronósticos a corto plazo generados por las metodologías expuestas en este trabajo pueden ser una herramienta valiosa para los decisores de salud pública para guiar en la asignación de recursos críticos necesarios para controlar la epidemia y

para responder a futuros brotes de enfermedades infecciosas.

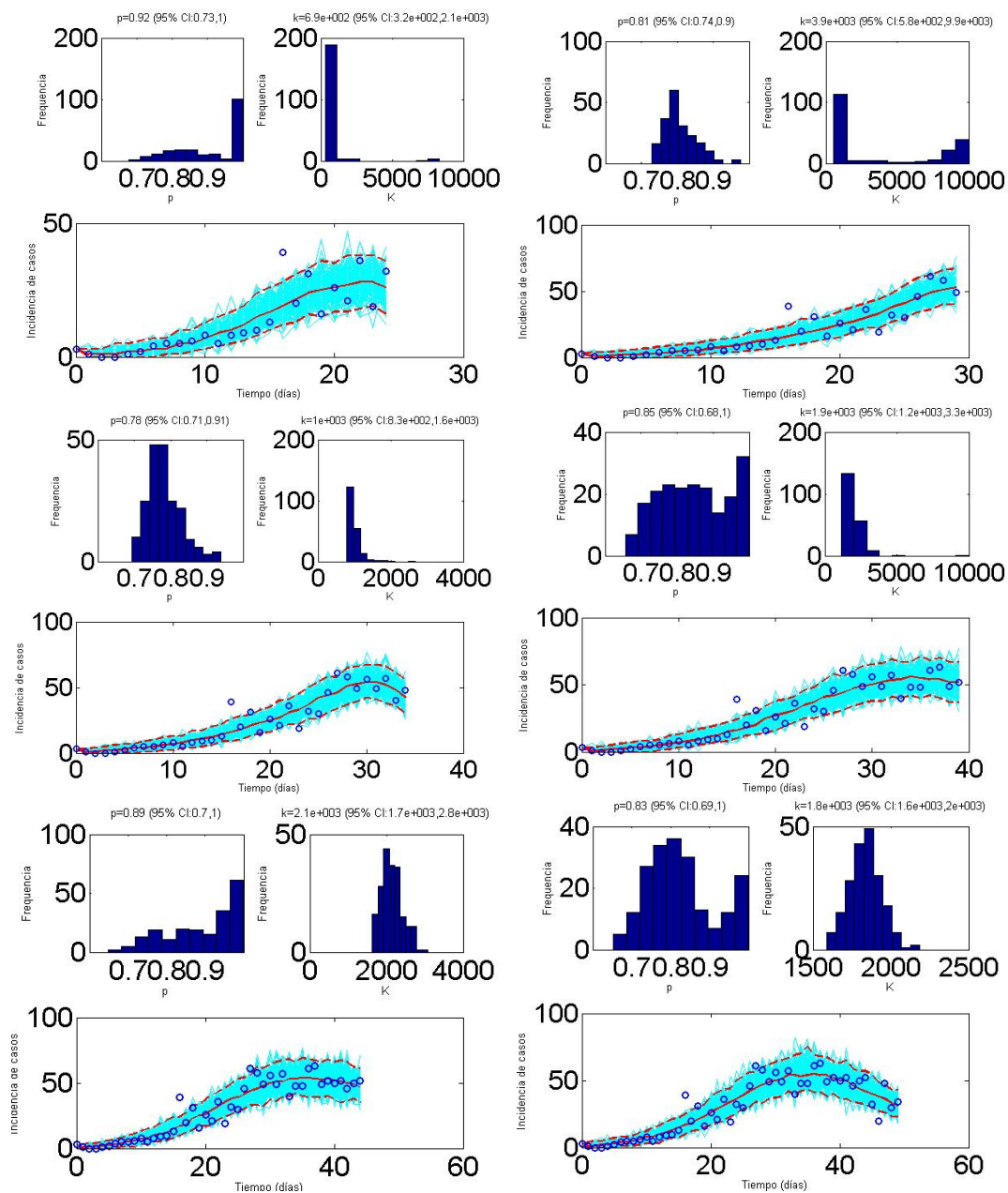
## Referencias

- [1] Chowell, G.: *Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts*. *Infect. Dis. Model.*, 2:379398, 2017.
- [2] Chowell, G., D. Hincapie-Palacio, J. Ospina, B. Pell, A. Tariq, S. Dahal, S. Moghadas, A. Smirnova, L. Simonsen y C. Viboud: *Using Phenomenological Models to Characterize Transmissibility and Forecast Patterns and Final Burden of Zika Epidemics*. *PLOS Currents Outbreaks*, May 2016.
- [3] Chowell, G., L. Sattenspiel, S. Bansald y C. Viboud: *Mathematical models to characterize early epidemic growth: A review*. *Physics of Life Reviews*, July 2016.
- [4] Chowell, G. y C. Viboud: *Is it growing exponentially fast? -Impact of assuming exponential growth for characterizing and forecasting epidemics with initial near-exponential growth dynamics*. *Infectious Disease Modelling*, 2016.
- [5] Chowell, G., C. Viboud, L. Simonsen y S. M. Moghadas: *Characterizing the reproduction number of epidemics with early subexponential growth dynamics*. *J. R. Soc. Interface*, 13(20160659), 2016.
- [6] Firat, M., E. Karaman, E. Basar y D. Narinc: *Bayesian Analysis for the Comparison of Nonlinear Regression Model Parameters: an Application to the Growth of Japanese Quail*. *Brazilian Journal of Poultry Science*, Special issue(Quails):19–26, 2015.
- [7] Hsieh, Y.H.: *Richards Model: A Simple Procedure for Real-time Prediction of outbreak Severity*. En Ma, Zhien, Jianhong Wu y Yicang Zhoue (editores): *Modeling and Dynamics of Infectious Diseases (Volumen 11)*, Contemporary Applied Mathematics (CAM), páginas 216–236. Higher Education Press, 2009, ISBN 978-981-4261-25-8. <http://mail.cmu.edu.tw/~hsieh/pdf/pub57.pdf>.
- [8] Hsieh, Y.H.: *Temporal course of 2014 Ebola virus disease (EVD) outbreak in West Africa elucidated through morbidity and mortality data: a tale of three countries*. *PLoS ONE*, 10(11):e0140810, 2015.
- [9] Hsieh, Y.H.: *Temporal patterns and geographic heterogeneity of Zika virus (ZIKV) outbreaks in French Polynesia and Central America*. *PeerJ*, (5):e3015, 2017.
- [10] Hsieh, Y.H., H. Arazoza y R. Lounes: *Temporal trends and regional variability of 2001-2002 multiwave DENV-3 Epidemic in Havana City: did Hurricane Michelle contribute to its severity?* *Tropical Medicine and International Health*, 18(7):830–838, July 2013.
- [11] Hsieh, Y.H., H.M. Huang y Y.C. Lan: *On Temporal Patterns and Circulation of Influenza Virus Strains in Taiwan, 2008-2014: Implications of 2009 pH1N1 Pandemic*. *PLoS ONE*, 11(5):e0154695, 2016.
- [12] Lindstrom, M.J. y D.M. Bates: *Nonlinear mixed-effects models for repeated measures data*. *Biometrics*, 46:673–687, 1990.
- [13] Ma, J., J. Dushoff, B. Bolker y D. Earn: *Estimating initial epidemic growth rate*. *Bull Math Biol*, 76:245–260, 2014.
- [14] Richards, F.J.: *A Flexible Growth Function for Empirical Use*. *Journal of Experimental Botany*, 10(29):290–301, 1959.
- [15] Roosa, K., Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. Hyman, P. Yan y G. Chowell: *Real-time forecasts of the 2019-nCoV epidemic in China from February 5th to February 24th, 2020*. *Infect. Dis. Model.*, 2020.
- [16] Roosa, K., Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. Hyman, P. Yan y G. Chowell: *Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, China: February 1323, 2020*. *Journal of Clinical Medicine*, 9:596, 2020.
- [17] Sebrango, C., L. Sanchez y Z. Shkedy: *Modeling dengue outbreak data using nonlinear mixed effects model*. *Investigación Operacional*, 38(3):252–258, 2017.
- [18] Smirnova, A. y G. Chowell: *A primer on stable parameter estimation and forecasting in epidemiology by a problem-oriented regularized least squares algorithm*. *Infectious Disease Modelling*, 2(2):268–275, 2017.
- [19] Tsoularis, A. y J Wallace: *Analysis of Logistic Growth Models*. *Mathematical Biosciences*, 179(1):21–55, July-August 2002.
- [20] Wang, X., J. Wu y Y. Yang: *Richards model revisited: validation by and application to infection dynamics*. *Journal of Theoretical Biology*, 313:12–19, 2012.

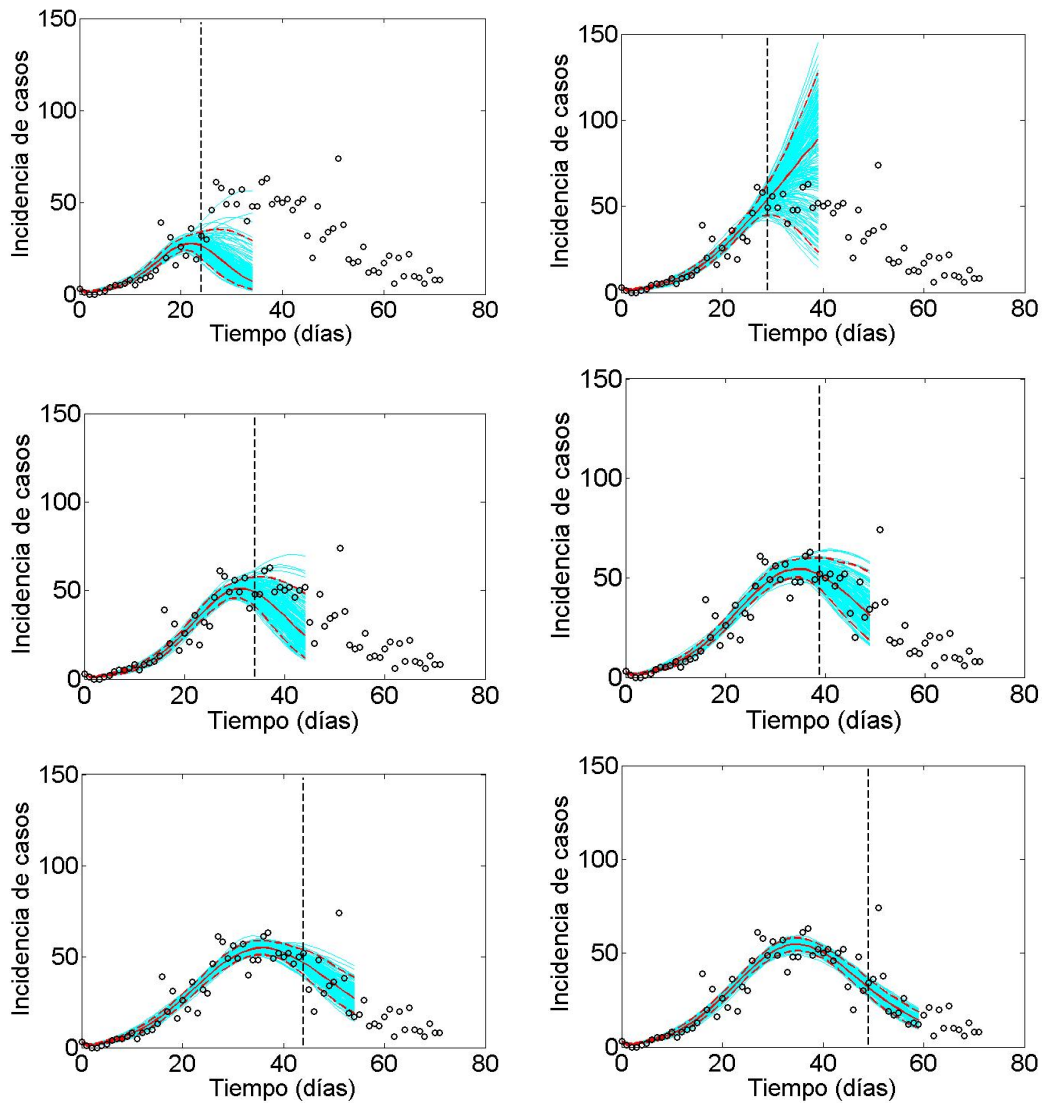


**Figura 1.** Estimación de los parámetros con incertidumbre cuantificada generada. Los histogramas muestran las distribuciones empíricas de las estimaciones de parámetros  $p$  y  $k$  utilizando 200 realizaciones bootstrap. El panel de abajo muestra el ajuste del MLG a los diferentes períodos. Los puntos azules son los datos diarios mientras que la línea roja continua corresponde al mejor ajuste del MLG a los datos. Las líneas azules corresponden a las 200 realizaciones de la curva epidémica asumiendo una estructura de error de Poisson. Las líneas rojas discontinuas corresponden a las bandas de confianza del 95 % alrededor del mejor ajuste.

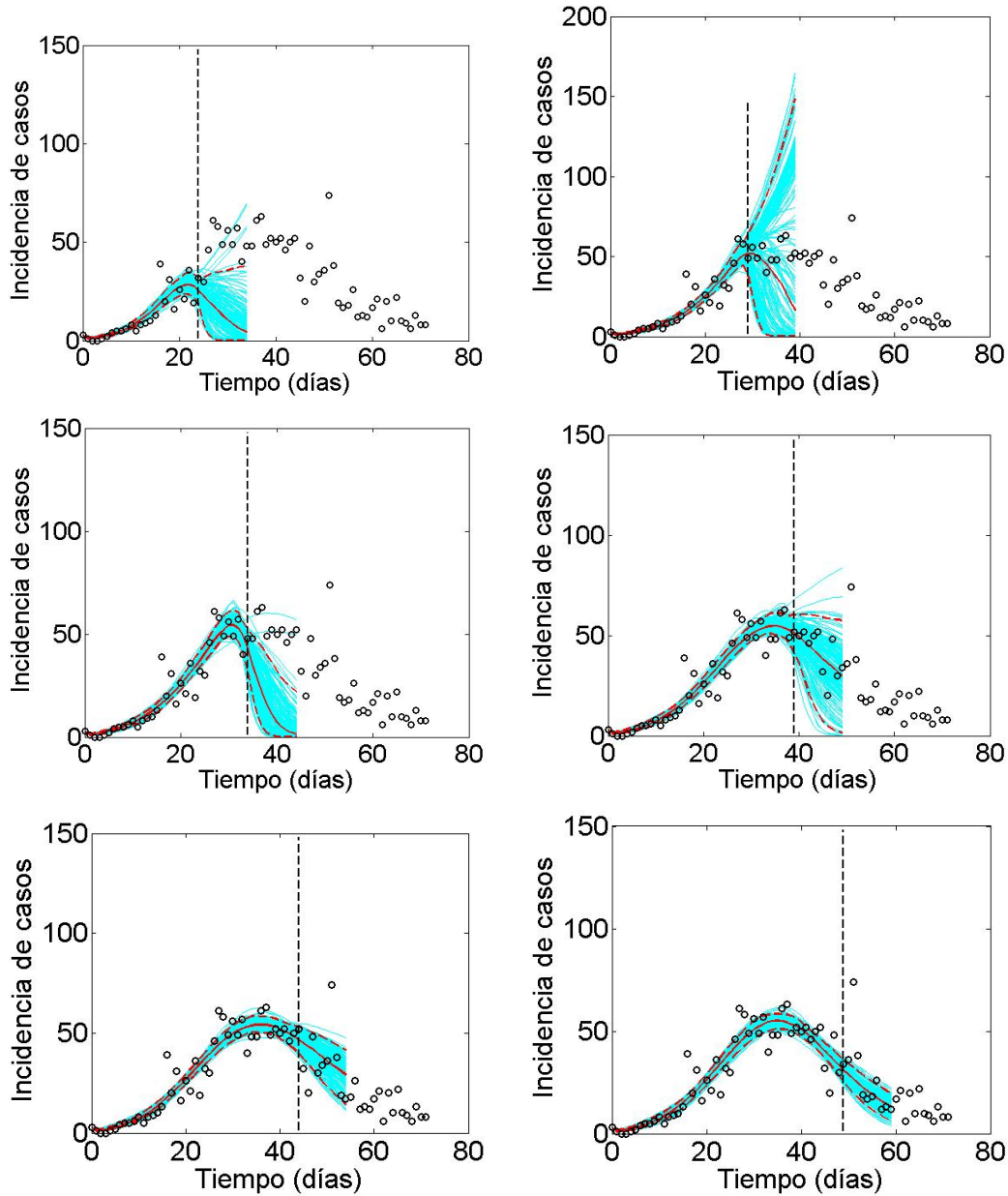




**Figura 2.** Ajustando el modelo logístico generalizado MRG a los primeros (25, 30, 35, 40, 45 y 50) días de la epidemia Covid-19 en Cuba. Estimación de los parámetros con incertidumbre cuantificada generada. Los histogramas muestran las distribuciones empíricas de las estimaciones de parámetros  $p$  y  $k$  utilizando 200 realizaciones bootstrap. El panel de abajo muestra el ajuste del MRG a los diferentes períodos. Los puntos azules son los datos diarios mientras que la línea roja continua corresponde al mejor ajuste del MRG a los datos. Las líneas azules corresponden a las 200 realizaciones de la curva epidémica asumiendo una estructura de error de Poisson. Las líneas rojas discontinuas corresponden a las bandas de confianza del 95% alrededor del mejor ajuste.



**Figura 3.** Pronósticos epidémicos de 10 días basados en el modelo de logístico generalizado (MLG) calibrado utilizando una cantidad creciente de datos epidémicos (círculos negros): (A) 25, (B) 30, (C) 35, (D) 40, (E) 45 y (F) 50 días epidémicos. La línea discontinua vertical indica el final del período de calibración. Se muestran la media (línea roja continua) y el IC del 95% (líneas rojas discontinuas) de los conjuntos de ajuste del modelo (curvas azules).



**Figura 4.** Pronósticos epidémicos de 10 días basados en el modelo de Richards generalizado (MRG) calibrado utilizando una cantidad creciente de datos epidémicos (círculos negros): (A) 25, (B) 30, (C) 35, (D) 40, (E) 45 y (F) 50 días epidémicos. La línea discontinua vertical indica el final del período de calibración. Se muestran la media (línea roja continua) y el IC del 95 % (líneas rojas discontinuas) de los conjuntos de ajuste del modelo (curvas azules).

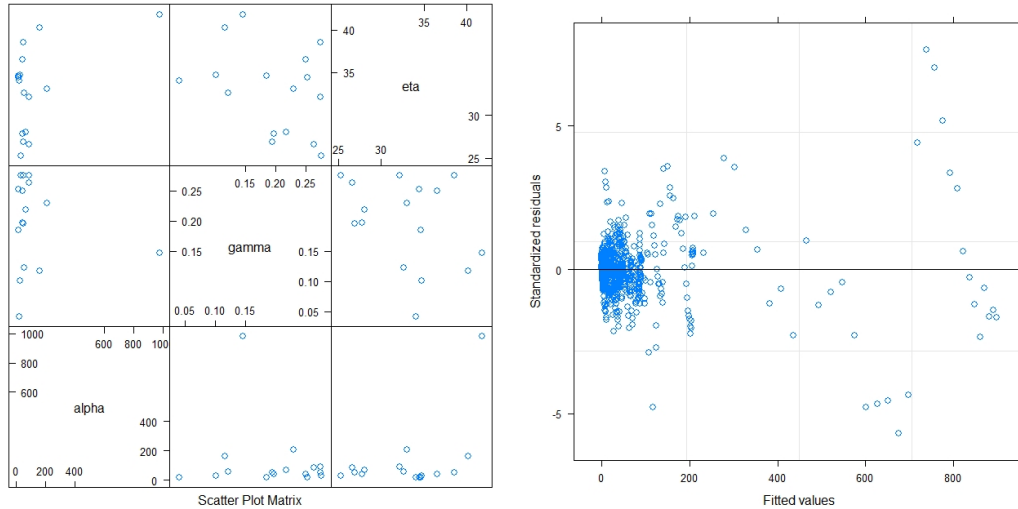


Figura 5. Gráfico de dispersión para las estimaciones de los efectos aleatorios del modelo con  $\alpha$ ,  $\eta$  y  $\gamma$  como efectos aleatorios (Panel izquierdo) y Gráfico de dispersión de los errores estandarizados (Panel derecho) para el modelo con  $\Psi$  bloque diagonal.

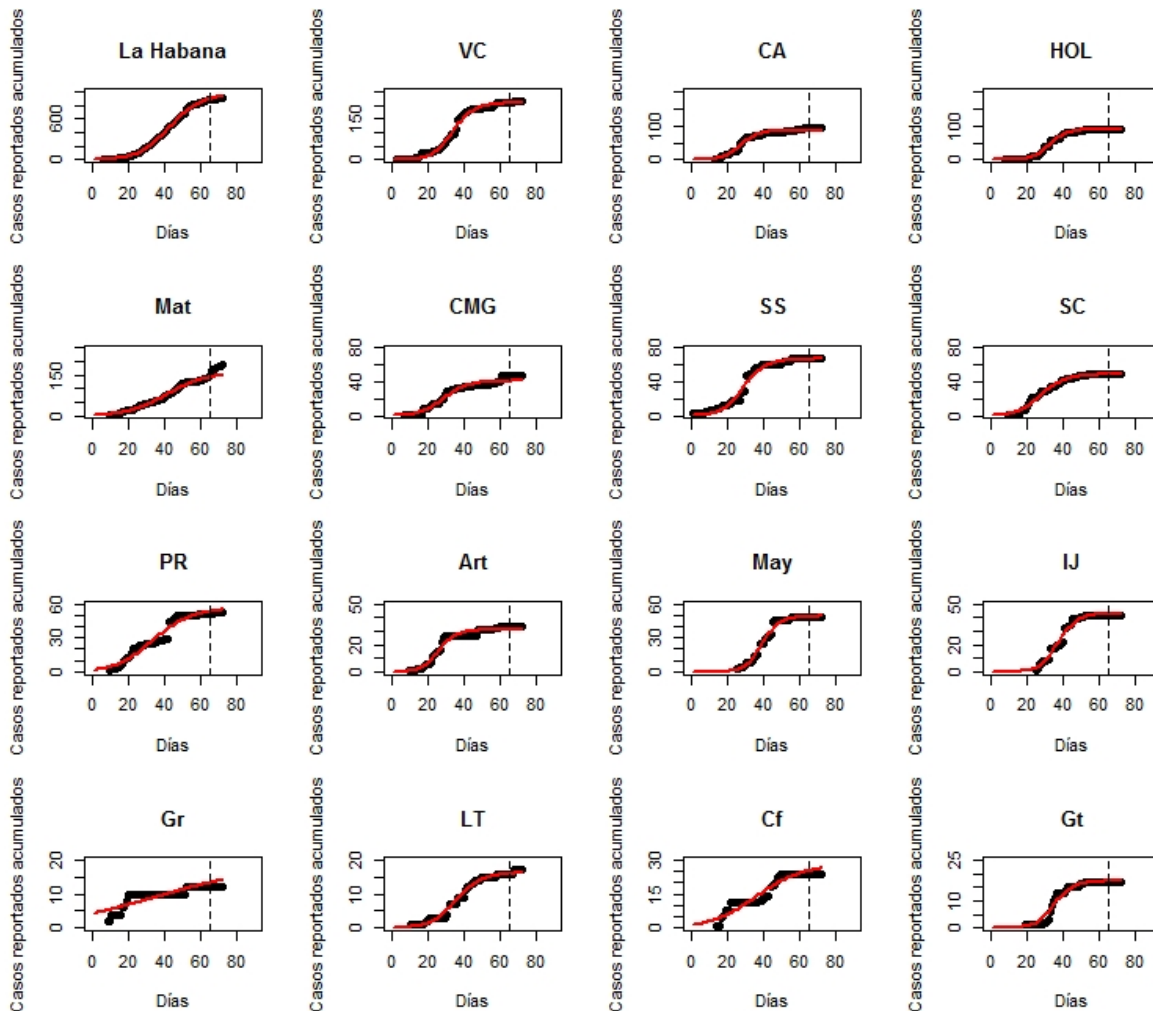


Figura 7. Ajuste y pronósticos (línea roja) con un horizonte de una semana del número de casos acumulados confirmados de todas las provincias de Cuba (puntos negros), para el modelo no lineal con efectos mixtos de mejor ajuste. La línea discontinua separa el período de calibración del de pronóstico.