

Modelado con Retardos Explícitos de la Propagación de COVID-19 en Argentina

Mariana Bergonzi
CIFASIS - CONICET, FCEIA - UNR
Rosario, Argentina
bergonzi@cifasis-conicet.gov.ar

Rodrigo Castro
ICC - CONICET, FCEN - UBA
Buenos Aires, Argentina
rcastro@dc.uba.ar

Ernesto Kofman
CIFASIS - CONICET, FCEIA - UNR
Rosario, Argentina
kofman@cifasis-conicet.gov.ar

Resumen—En este artículo se presenta un nuevo modelo compartimental de tiempo discreto de la propagación de COVID-19. Su particularidad es que tiene en cuenta de manera explícita los retardos correspondientes a las distintas etapas de la enfermedad y su diagnóstico y que permite distinguir la presencia de casos importados.

Además de desarrollar las ecuaciones del modelo, se describe un mecanismo de ajuste automático de sus parámetros usando datos oficiales de la propagación del virus en Argentina. El resultado de esta parametrización refleja consistentemente el comportamiento del virus respecto de sus tiempos característicos: latencia, período infeccioso, reporte de casos (confirmados y fallecidos) y además permite detectar automáticamente cambios en el número reproductivo y en el factor de mortalidad.

Se analiza además la capacidad de predicción del modelo y se muestran resultados de simulación suponiendo distintos escenarios a futuro.

I. INTRODUCCIÓN

Dentro de las distintas alternativas para describir modelos epidemiológicos, los de tipo SIR, SEIR y SEIRD se encuentran entre los más difundidos [1]. Estos modelos en general se formulan en tiempo continuo (ecuaciones diferenciales) aunque hay variantes de tiempo discreto [2]. Para el caso particular del COVID-19, hay ya publicados algunos trabajos donde buscan ajustar modelos de este tipo tanto en tiempo continuo [3], [4], [5] como en tiempo discreto [6], [7].

En el caso del COVID-19, una característica notable del virus es su tiempo de incubación prolongado que demora primero el inicio del período de contagio y luego el momento de la detección. Por este motivo, parece apropiado incluir explícitamente en el modelo la presencia de dichos *retardos*. Además, en una etapa temprana de propagación del virus en una sociedad, son muy frecuentes y determinantes los casos importados. Esto es particularmente cierto en Argentina, donde más de un mes después del diagnóstico del primer caso, la mitad de los casos acumulados detectados seguían siendo importados.

Dado que la incorporación explícita de retardos y casos importados reportados diariamente complica el planteo por primeros principios de un sistema de ecuaciones diferenciales (implica utilizar ecuaciones diferenciales con retardo, cuya resolución numérica puede ser compleja [8]), en este trabajo proponemos realizar un modelo de tiempo discreto con paso de un día y que, a diferencia de los modelos tradicionales de tiempo discreto [1], tenga en cuenta los tiempos de incubación, detección, recuperación y muerte. Además, como mencionamos, en este modelo incorporamos los casos importados como dato de entrada.

En este artículo desarrollamos dicho modelo y mostramos los resultados de ajustar mediante mínimos cuadrados todos sus parámetros con datos públicos brindados por el Ministerio de Salud de Argentina. Como se verá, los parámetros que minimizan el error cuadrático medio entre los datos del Ministerio y las predicciones del modelo coinciden con información conocida sobre el virus en nuestro país (tiempo de incubación, período de contagio, tiempos de recuperación, muerte, detección y reporte).

Una vez ajustados los parámetros, estudiamos la capacidad del modelo para predecir la evolución de contagios y muertes a mediano plazo y, finalmente, utilizamos el modelo para analizar posibles escenarios futuros.

El artículo está organizado de la siguiente forma. En la Sección II se introducen los conceptos básicos sobre los modelos de propagación de epidemias. Luego, en la Sección III se presentan las ecuaciones del modelo y se describe el ajuste de sus parámetros. A continuación, en la Sección IV se analiza la capacidad de predicción del modelo y se estudian distintos escenarios a futuro y finalmente en la Sección V se presentan las conclusiones.

II. ANTECEDENTES

II-A. Modelos SIR, SEIR y SEIRD

Existen muchos tipos de modelos matemáticos para representar, analizar y predecir los comportamientos y efectos producidos por una epidemia en una población determinada [1], [9]. Los mismos pueden ser determinísticos o estocásticos, de tiempo continuo o discreto, de compartimientos u orientados a agentes, etc., dependiendo de las características que más nos interese modelizar.

Dentro de las distintas alternativas para describir modelos epidemiológicos, los modelos de compartimientos tipo SIR, SEIR y SEIRD se encuentran entre los más difundidos. El origen de estos se remonta al trabajo de Kermack y A.G. McKendrick en el año 1927 [10] y consiste en clasificar la población en tres grupos (o *compartimientos*) diferentes:

- $S(t)$: Población *susceptible* al virus en tiempo t . (Individuos que no están infectados, ni tienen inmunidad al virus).
- $I(t)$: Población *infectada* en tiempo t . (Individuos que están infectados y pueden contagiar a los susceptibles).
- $R(t)$: Población *recuperada* en tiempo t . (Individuos que ya fueron *removidos* de la dinámica, ya sea porque se recuperaron adquiriendo inmunidad al virus, o bien murieron).

Cada individuo de la población estudiada pertenece a sólo uno de estos tres grupos y puede moverse de uno a otro. Una de las hipótesis es que los individuos recuperados adquieren inmunidad de por vida al virus, por lo que los movimientos permitidos pueden ser: Que un individuo de S pueda infectarse al estar en contacto con uno del grupo I , pasando a formar parte de este grupo. Uno del grupo I termina pasando al grupo R al recuperarse o morir. Entonces, el número de individuos en cada grupo cambia a través del tiempo. La cantidad total de individuos en la población se asume constante siendo $S(t) + I(t) + R(t) = N$.

Existen otros modelos matemáticos basados en el modelo SIR, a los que se le agregan otros compartimientos o características adicionales: En el modelo SEIR se agrega a las ya mencionadas la categoría de población *Expuesta* $E(t)$, que son los individuos que están incubando el virus pero todavía no tienen capacidad de contagio. Por su parte, el modelo SEIRD agrega al último la categoría $D(t)$ correspondiente a la población *Muerta* hasta el instante t .

II-A1. Modelos de tiempo continuo: En poblaciones grandes las variables pueden ser tomadas como

continuas y la relación entre los tres grupos se puede representar con un sistema de ecuaciones diferenciales ordinarias:

$$\begin{aligned}\frac{dS}{dt} &= -\beta \cdot S \cdot I \\ \frac{dI}{dt} &= \beta \cdot S \cdot I - \gamma \cdot I \\ \frac{dR}{dt} &= \gamma \cdot I\end{aligned}$$

En donde β y γ son parámetros que representan las tasas de infección y de recuperación respectivamente. De la relación de estos parámetros surge un indicador fundamental: R_0 , el número de reproducción básico. Este factor indica a cuantas personas en promedio contagia un infectado durante su tránsito por la enfermedad.

II-A2. Modelos en tiempo discreto: Existen también modelos en los que se discretiza el tiempo, manteniendo la misma clasificación y dinámica del modelo SIR [2], [1].

II-A3. Modelos con retardos: A los modelos mencionados, se les puede agregar retardos para modelar explícitamente los comportamientos que se manifiestan al cabo de un período de tiempo conocido [1]. Un ejemplo puede ser el *tiempo de latencia* de una enfermedad, que es el tiempo desde que el individuo se expone al virus hasta que comienza a contagiar.

II-B. Modelos para COVID-19

Para el caso particular del COVID-19 hay ya publicados múltiples trabajos con modelos compartimentales de diversa naturaleza. Entre los de tiempo continuo se pueden mencionar a [3], [4], [5] entre otros. Entre los de tiempo discreto, se pueden mencionar a [6], [7].

Respecto a la incorporación de retardos, hay también un trabajo de tiempo continuo [11], donde se propone una estructura tipo SEIR pero se plantea una pequeña diferencia respecto a las categorías. Aquí, la categoría I representa a *todos* los infectados (detectados y no detectados) que ya contagian, pero aun no tienen síntomas. Además, R representa a aquellos infectados detectados (la R es por *Reportados*) y U a los infectados *no detectados*.

En este trabajo se plantean dos opciones diferentes para representar el período de latencia. Por un lado mediante tasas para representar los flujos entre compartimientos y por otro, representando explícitamente el período de latencia mediante un retardo y obteniendo una *Ecuación Diferencial con Retardos*.

III. FORMULACIÓN DEL MODELO

En esta sección presentamos el modelo propuesto.

III-A. Ecuaciones del Modelo

Siguiendo la idea de los modelos SEIRD, plantearemos las siguientes variables de estado:

- $S(t)$: Población susceptible al virus al tiempo t .
- $E(t)$: Población expuesta al virus que todavía no contagia al tiempo t .
- $I(t)$: Población infectada con capacidad de contagio.
- $R(t)$: Población removida del modelo porque ya no contagia.
- $D(t)$: Población muerta a causa de la infección.

Luego, para incorporar explícitamente el efecto de los retardos, tendremos también como variables de estado al número de *nuevos expuestos en los T pasos anteriores* (donde T es el máximo retardo considerado):

$$N^E(t), N^E(t-1), \dots, N^E(t-T)$$

Para tener en cuenta el efecto de la llegada de casos importados, consideraremos una señal $U(t)$ que asumiremos conocida y que representará el número de casos importados que se contagiaron al instante t (estando en el exterior) y que serán detectados en el país. Notar que no son necesariamente todos los casos. Utilizaremos, además, los siguientes parámetros (que luego ajustaremos según se describe en la Sección III-B) y que consideraremos como valores promedio:

- g : Factor de diagnóstico. Esto es, fracción de los casos positivos que son eventualmente diagnosticados.
- $d(t)$: Fracción de casos positivos diagnosticados que finalmente mueren (que supondremos puede cambiar con el tiempo).
- $r_0(t)$: Factor de contagio por individuo y por día (que supondremos puede cambiar con el tiempo).
- τ_I : Tiempo de latencia. Esto es, tiempo desde que el individuo se expone hasta que se torna contagioso.
- τ_R : Tiempo de fin de contagio. Esto es, tiempo desde que un individuo se expone hasta que deja de contagiar. Notar que $\tau_R - \tau_I = \tau_{\text{Inf}}$ es la duración del período infeccioso.
- τ_D : Tiempo de diagnóstico. Desde la exposición hasta que es diagnosticado y reportado.
- τ_M : Tiempo de muerte. Desde la exposición hasta que se muere, en caso que tenga ese desenlace.

El factor de contagio por individuo y por día, será calculado como $r_0(t) = R_0(t)/(\tau_R - \tau_I)$, donde R_0 es el número reproductivo básico (cuantos contagios produce un contagio individual).

En base a estas variables y parámetros, el modelo tiene la siguiente dinámica:

$$\begin{aligned} N^E(t+1) &= r_0 \cdot I(t) \cdot S(t) + U(t+1)/g \\ S(t+1) &= S(t) - r_0 \cdot I(t) \cdot S(t) \\ E(t+1) &= E(t) + N^E(t+1) - N^E(t+1 - \tau_I) \\ I(t+1) &= I(t) + N^E(t+1 - \tau_I) - N^E(t+1 - \tau_R) \\ R(t+1) &= R(t) + N^E(t+1 - \tau_R) \end{aligned}$$

Además, para ajustar posteriormente los parámetros y realizar proyecciones, se calculan las variables que son habitualmente reportadas:

$$\begin{aligned} D(t+1) &= D(t) + (d \cdot g) \cdot N^E(t+1 - \tau_M) \\ I_D(t+1) &= I_D(t) + g \cdot N^E(t+1 - \tau_D) \end{aligned}$$

que representan el número total de muertes y el número total de casos diagnosticados y reportados.

El modelo se basa en las siguientes hipótesis:

- Los nuevos casos son proporcionales al número de personas infectadas y al número de personas susceptibles más el número de casos importados.
- La fracción de casos diagnosticados g es constante y única tanto para casos locales como importados.

Puede verse además que si se mantiene aproximadamente constante $S(t)$ (esto es así salvo que el número de infecciones sea muy grande y baje significativamente la población susceptible, lo que aún no ocurre), el modelo resulta lineal. Si además las condiciones iniciales son nulas y la dinámica aparece por la entrada de casos importados reportados $U(t)$, el valor de g no influye en el número de muertes ni en el número de casos reportados y detectados. El motivo es que g aparece dividiendo la entrada y luego aparece multiplicando dichas salidas siendo un modelo lineal.

Esta característica tiene la ventaja que permite inferir la dinámica independientemente del factor de detección. De todas formas, una limitación de este enfoque es que no conocer el factor g no permite inferir el número real de gente infectada, sino sólo una proporción de esta.

III-B. Ajuste de Parámetros

Para ajustar los parámetros, implementamos el modelo anterior en Octave utilizando como entrada los datos de casos importados en Argentina en el período

mencionado. Se fijó como tiempo inicial $t = 0$ el día 20/2/2020, ya que los primeros casos reportados y la primera muerte datan de comienzos de marzo lo que lleva el tiempo de exposición del primer caso alrededor de dicha fecha.

Tomamos como salida del modelo el número de muertes $D(t)$ y el número de casos totales detectados $I_D(t)$. Estos números fueron comparados con los datos correspondientes (de muertes y reportes de casos) desde el 20/2/2020 hasta el 25/4/2020. Todos los datos fueron obtenidos de los informes diarios del Ministerio de Salud de la Nación <https://www.argentina.gob.ar/coronavirus/informe-diario>.

Para cada juego de valores enteros de los retardos $\tau_I, \tau_D, \tau_M, \tau_R$ se ajustó por mínimos cuadrados los valores de $R_0(t)$ (tasa de reproducción) y $d(t)$ (fracción de muertes por caso detectado y reportado), y luego se buscó que juego de valores de retardo minimizaba el costo resultante. Para el ajuste por mínimos cuadrados se utilizó un peso en cada salida inversamente proporcional a la raíz cuadrada de la norma de cada señal (de manera que el ajuste sea acorde a los valores relativos de las señales). Esto último busca minimizar la norma del error relativo en cada salida.

Se optimizó también sobre dos instantes de cambio para los parámetros $R_0(t)$ y $d(t)$ dentro del período de ajuste. El resultado de toda la optimización resultó en las siguientes expresiones para las tasas de reproducción y mortalidad:

$$R_0(t) = \begin{cases} 2.267 & \text{si } t < 17/3 \\ 0.886 & \text{si } 17/3 \leq t < 3/4 \\ 1.104 & \text{si } t \geq 3/4 \end{cases}$$

$$d(t) = \begin{cases} 0.054 & \text{si } t < 17/3 \\ 0.061 & \text{si } 17/3 \leq t < 3/4 \\ 0.072 & \text{si } t \geq 3/4 \end{cases}$$

Cabe aclarar que en todos los casos el algoritmo de mínimos cuadrados se ejecutó con parámetros iniciales de 1 para $R_0(t)$ y 0.05 para $d(t)$. Se verificó además convergencia a este mínimo desde distintos parámetros iniciales.

Dichos valores de los parámetros se obtuvieron para retardos $\tau_I = 5$, $\tau_D = 11$, $\tau_R = 12$, y $\tau_M = 17$. Estos valores de retardo que minimizan el costo cuadrático son consistentes con lo que se sabe del virus: 5 días hasta que el paciente se torna contagioso, 11 días es un tiempo razonable hasta la detección y reporte, hasta el día número 12 contagia (esto es coherente pensando

que a partir de la detección se aísla) y 17 días hasta la recuperación o muerte. Asimismo, se destaca que el instante de tiempo t_1 que se obtiene para que el error sea mínimo resultó casi coincidente con la aplicación de la cuarentena estricta (t_1 se corresponde con el 17/03, mientras que la cuarentena se indicó el 19/03). Mientras que el tiempo t_2 obtenido se corresponde al 03/04, cuando la población comenzó a relajar las condiciones estrictas de confinamiento.

En el ajuste, el error medio relativo entre datos medidos y simulados es $(x_{\text{data}} - x_{\text{sim}})/x_{\text{data}} = 0.0613$ (6.13%). Las Figuras 1-2 comparan los resultados de simulación con los datos correspondientes al número de muertes, número de casos detectados totales y por día.

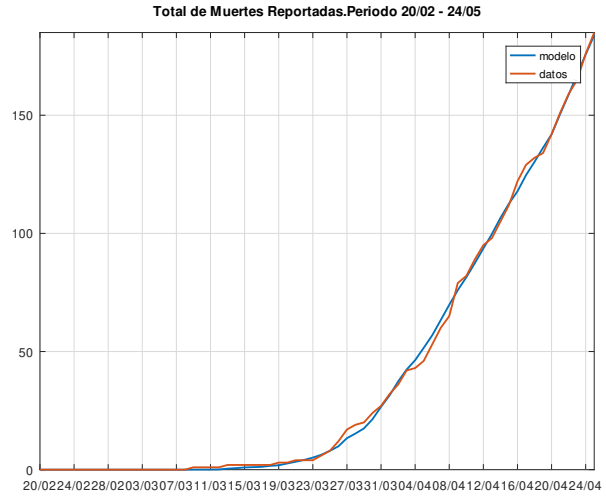


Figura 1. Número total de muertes (datos y modelo)

IV. EXPERIMENTOS Y RESULTADOS

En esta sección mostraremos el uso del modelo desarrollado para analizar la evolución de los factores de contagio y de mortalidad en períodos pasados, por un lado, y para proyectar la evolución de las curvas a corto y mediano plazo.

Para esto, realizaremos el ajuste de los parámetros $R_0(t)$ y $d(t)$ en un período más extenso, utilizando datos desde el 20/2/2020 al 04/6/2020. Consideraremos que estos parámetros pueden cambiar cada 14 días. Manteniendo los parámetros de retardo del experimento anterior ($\tau_I = 5$, $\tau_D = 11$, $\tau_R = 12$ y $\tau_M = 17$) se obtienen entonces las siguientes trayectorias de $R_0(t)$ y $d(t)$:

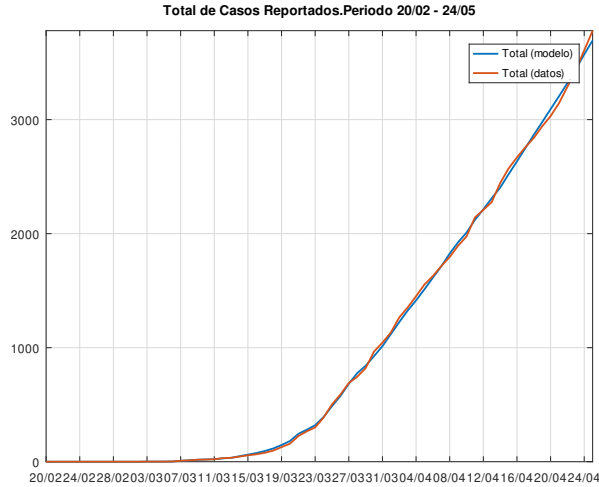


Figura 2. Número total de casos detectados (datos y modelo)

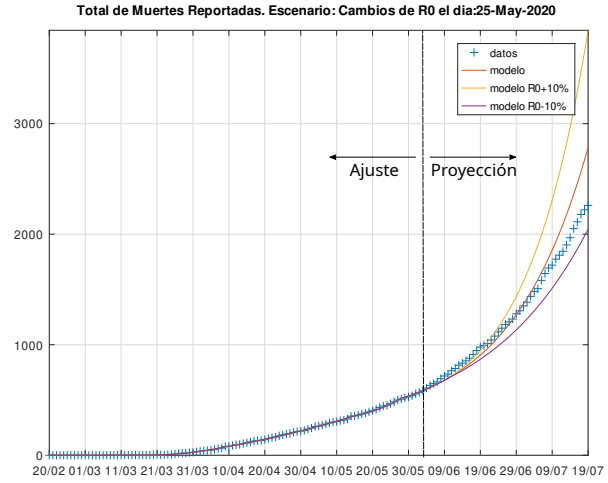


Figura 3. Número total de muertes (datos y modelo)

$$R_0(t) = \begin{cases} 2.396 & \text{si } t < 16/3 \\ /0.898 & \text{si } 16/3 \leq t < 29/3 \\ 1.051 & \text{si } 30/3 \leq t < 12/4 \\ 1.078 & \text{si } 13/4 \leq t < 26/4 \\ 2.042 & \text{si } 27/4 \leq t < 10/5 \\ 1.523 & \text{si } t \geq 11/5 \end{cases} \quad (1)$$

$$d(t) = \begin{cases} 0.054 & \text{si } t < 16/3 \\ 0.061 & \text{si } 16/3 \leq t < 29/3 \\ 0.064 & \text{si } 30/3 \leq t < 12/4 \\ 0.068 & \text{si } 13/4 \leq t < 26/4 \\ 0.031 & \text{si } 27/4 \leq t < 10/5 \\ 0.018 & \text{si } t \geq 11/5 \end{cases} \quad (2)$$

En el ajuste, el error medio relativo entre datos medidos y simulados resulta $\|x_{\text{data}} - x_{\text{sim}}\|/\|x_{\text{data}}\| = 0.015$ (1.5%).

Las Figuras 3-6 comparan los resultados de simulación con los datos correspondientes al número de muertes, número de casos detectados totales y por día. En este caso, extendimos el tiempo final de simulación para generar una proyección más allá de los datos utilizados para el ajuste. Consideramos inicialmente que para todo tiempo posterior al 12/5 los parámetros $R_0(t)$ y $d(t)$ permanecen constantes e iguales al último valor ajustado. Además incluimos dos escenarios considerando una variación de $\pm 10\%$ de estos parámetros a partir del 25/5, fecha desde la cual lo que ocurre no es observable en los datos usados para el ajuste debido a los retardos.

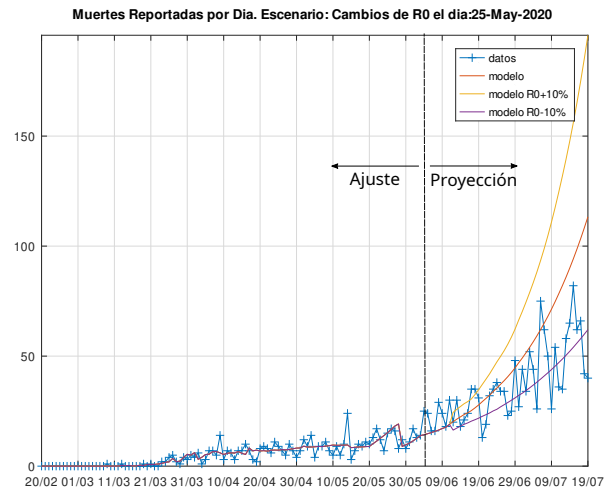


Figura 4. Número diario de muertes (datos y modelo)

Como vemos, el ajuste del modelo nos permite en primer lugar sacar conclusiones sobre la curva de datos pasados. Las Ecs.(1)-(2) nos muestran la evolución del factor de contagios y los efectos de la cuarentena y su posterior flexibilización, así como la disminución de la mortalidad por caso que refleja la mayor propagación del virus entre la población más joven.

Por otro lado, las proyecciones a partir del 4/6 (fecha final de los datos usados para el ajuste) muestran primero la posibilidad de usar el modelo para estimar a futuro la propagación del virus ante distintos escenarios. Por otro lado, permite analizar el efecto a futuro del aumento o disminución del distanciamiento social reflejado en el factor R_0 .

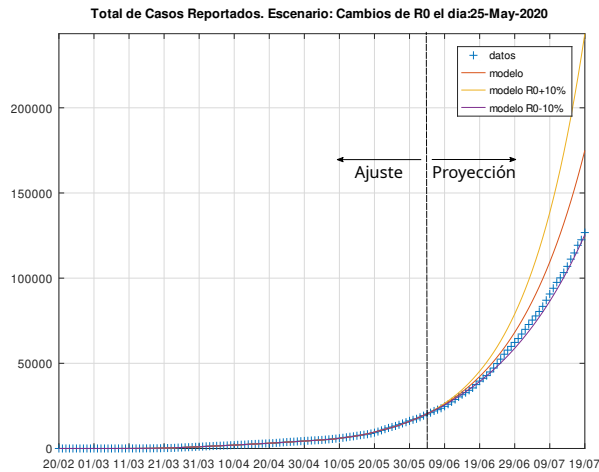


Figura 5. Número total de casos detectados (datos y modelo)

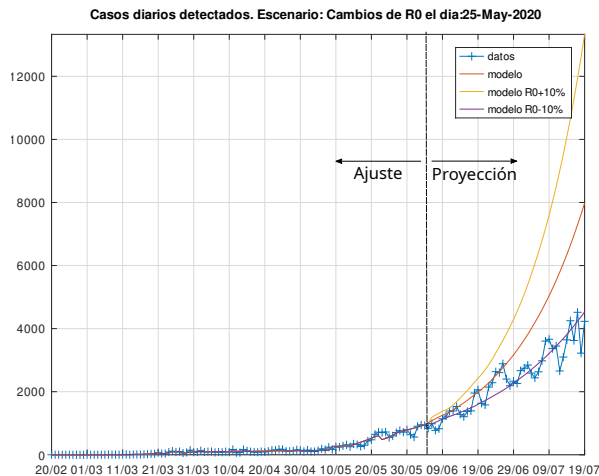


Figura 6. Número diario de casos detectados (datos y modelo)

V. CONCLUSIONES

Se presentó un modelo muy simple de la propagación de COVID-19 basado en primeros principios que no requiere hipótesis, desarrollos ni interpretaciones complejas. Los parámetros del mismo son transparentes y representan explícitamente los fenómenos característicos esenciales. Además, el modelo distingue los *casos importados*, lo que evita la sobre-estimación de contagios.

Otro aspecto a destacar de la metodología es que no se presuponen valores para los parámetros, sino que los mismos se obtienen mediante ajuste por mínimos cuadrados respecto de las mediciones reales. Es decir, utilizando sólo los datos reportados sobre el número de casos y muertes, el modelo permite inferir parámetros

biológicos, de comportamiento social, e incluso administrativos que influyen en la evolución y propagación del virus, permitiendo un análisis de los mismos.

Al tener parámetros explícitos y de interpretación directa es posible estudiar escenarios mediante su modificación que serían más engorrosos de representar en modelos clásicos de tiempo continuo. Por ejemplo, el efecto de una estrategia de seguimiento y aislamiento de contactos estrechos de casos confirmados (durante los N días previos a su detección) se puede capturar simplemente con una reducción del parámetro τ_R .

Asimismo, en lo que respecta a estrategias de control (intervenciones no farmacológicas) contar con un modelo con retardos explícitos entre actuación y medición es fundamental para poder estudiar el desempeño a lazo cerrado.

REFERENCIAS

- [1] F. Brauer, C. Castillo-Chavez, and C. Castillo-Chavez, *Mathematical models in population biology and epidemiology*. Springer, 2012, vol. 2.
- [2] L. J. Allen, "Some discrete-time si, sir, and sis epidemic models," *Mathematical biosciences*, vol. 124, no. 1, pp. 83–105, 1994.
- [3] J. T. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, P. M. de Salazar, B. J. Cowling, M. Lipsitch, and G. M. Leung, "Estimating clinical severity of covid-19 from the transmission dynamics in wuhan, china," *Nature Medicine*, pp. 1–5, 2020.
- [4] W. C. Roda, M. B. Varughese, D. Han, and M. Y. Li, "Why is it difficult to accurately predict the covid-19 epidemic?" *Infectious Disease Modelling*, 2020.
- [5] Y. Fang, Y. Nie, and M. Penny, "Transmission dynamics of the covid-19 outbreak and effectiveness of government interventions: A data-driven analysis," *Journal of medical virology*, 2020.
- [6] C. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos, "Data-based analysis, modelling and forecasting of the covid-19 outbreak," *PloS one*, vol. 15, no. 3, p. e0230405, 2020.
- [7] J. Dehning, J. Zierenberg, F. P. Spitzner, M. Wibral, J. P. Neto, M. Wilczek, and V. Priesemann, "Inferring change points in the spread of covid-19 reveals the effectiveness of interventions," *Science*, 2020.
- [8] R. Castro, E. Kofman, and F. E. Cellier, "Quantization-based integration methods for delay-differential equations," *Simulation Modelling Practice and Theory*, vol. 19, no. 1, pp. 314–336, 2011.
- [9] E. Vynnycky and R. White, *An introduction to infectious disease modelling*. OUP oxford, 2010.
- [10] W. O. Kermack and 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*, vol. 115, no. 772, pp. 700–721, 1927.
- [11] Z. Liu, P. Magal, O. Seydi, and G. Webb, "A covid-19 epidemic model with latency period," *Infectious Disease Modelling*, 2020.