

Reporte técnico. Grupo de análisis estadístico y modelos COVID-19, UNAM. Mayo 4, 2020.

## Estudio de dinámica global y local de propagación de COVID-19 con apoyo de modelos estocásticos y en datos globales.

Estephanie Escobar Sánchez<sup>1</sup>, José Guiral Barajas<sup>1</sup>, Carlos Ignacio Herrera-Nolasco<sup>1</sup>, Marco Arieli Herrera-Valdez<sup>1,✉</sup>, Sergio Iván López-Ortega<sup>1</sup>, Eugenia O'Reilly-Regueiro<sup>4</sup>,

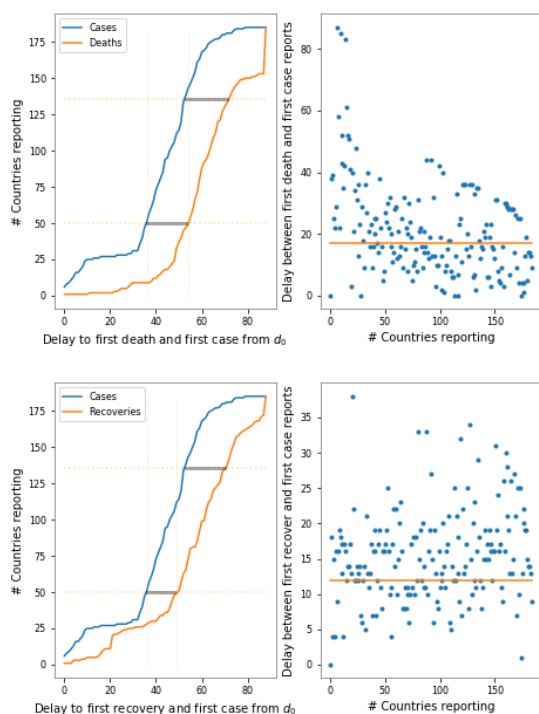
<sup>1</sup> Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México

<sup>2</sup> Instituto de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México

✉ marcoh@ciencias.unam.mx

Los datos presentados aquí fueron obtenidos de la base de datos ([https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\\_covid\\_19\\_data/csse\\_covid\\_19\\_time\\_series/](https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/))

### Inferencia de características del curso temporal de COVID-19 a partir de medidas macroscópicas de reportes de casos, recuperaciones y muertes



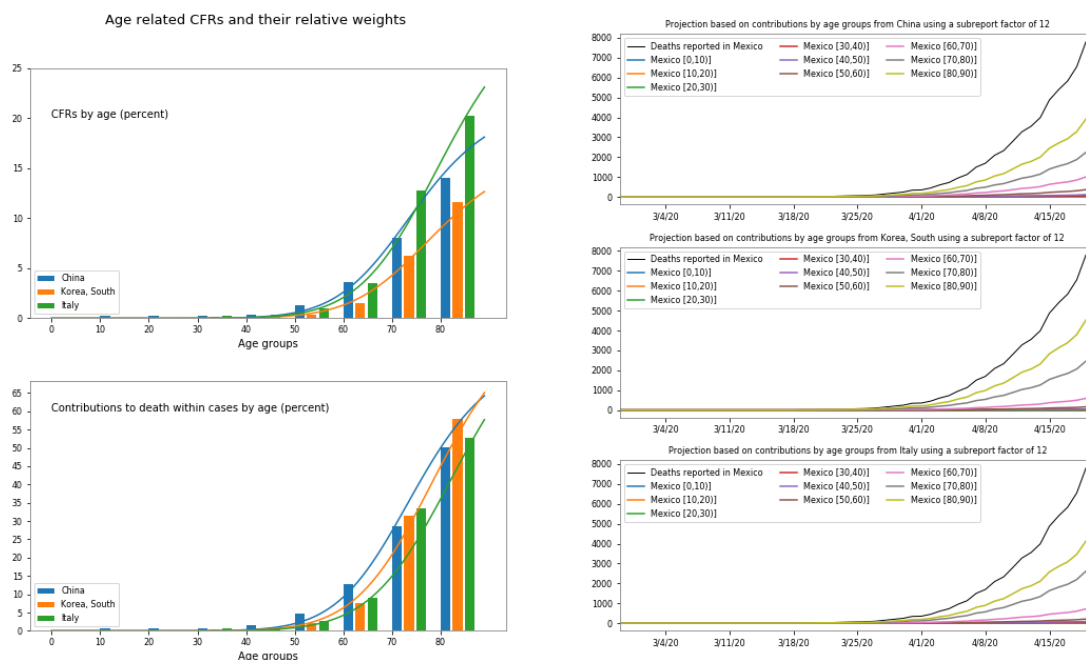
**Figura 1:** Retrasos entre primeros reportes reportes de casos y muertes (arriba), y entre casos y recuperaciones (abajo).

**Retrasos entre comienzo de síntomas y recuperación o muerte a partir de los primeros reportes de caso en el mundo** ¿Es posible observar en los reportes de casos hechos a nivel mundial indicaciones del tiempo que tarda una persona en morir por complicaciones de COVID-19 a partir comienza a tener síntomas? Es decir, ¿es posible estimar el retraso en entre la muerte y la manifestación de síntomas desde una perspectiva macroscópica de los casos y los decesos por caso en distintos países? ¿Se puede hacer el mismo tipo de inferencia tomando en cuenta retrasos entre recuperaciones y confirmaciones de caso?

La respuesta está en analizar el retraso entre primer reporte de casos y el primer reporte de muertes en distintos países. La mayoría de los países que primero reportaron casos también reportó muertes (sin retraso). Sin embargo, a medida que pasaron las semanas, se fue observando una tendencia que arroja un retraso promedio de 18 días entre primer caso confirmado y primera muerte en varios países (Fig. 1).

Los retrasos se pueden inferir utilizando distintos métodos, entre los cuales está simplemente medir

la distancia horizontal entre las gráficas acumuladas del número de países que han reportado primer caso o muerte, respectivamente (Fig. 1, en medio izquierda, y arriba a la derecha). El retraso correspondiente entre entre primeros reportes de casos y primera recuperación, es aproximadamente 12 días (Fig. 1 paneles de abajo).



**Figura 2:** (Izquierda) Contribuciones de distintos grupos de edad a la fatalidad por casos en China, Korea del Sur, e Italia. (Derecha) Estimaciones de fatalidad para México tomando en cuenta grupos de edad, con los datos de China, Korea del Sur, e Italia, calculado hasta el 11 de abril de 2020, y usando un factor de ajuste por subreporte igual a 12.

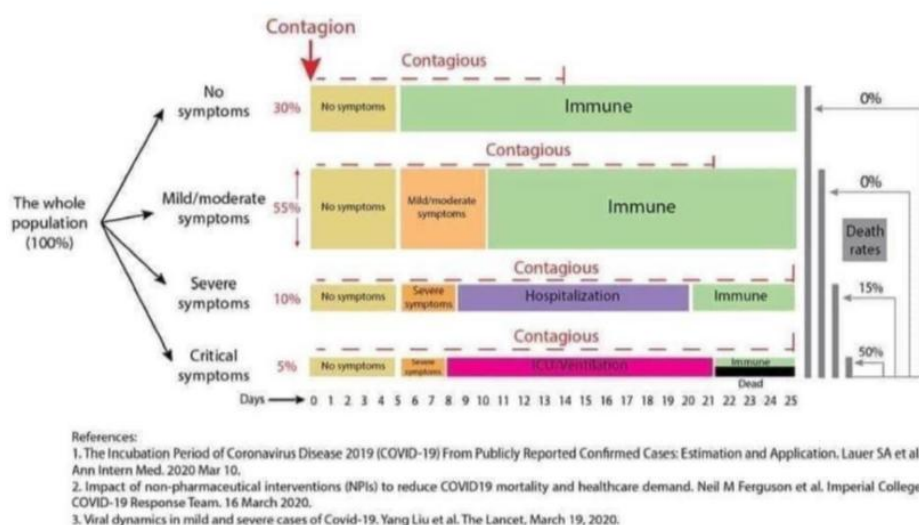
**Contribución de grupos de riesgo por edad a la dinámica de fallecimientos por caso.** La fatalidad en casos confirmados depende de factores que incluyen la calidad de los servicios de salud, y el acceso a dichos servicios, entre otros. Sin embargo, es posible observar algunas similitudes macroscópicas en el comportamiento de las defunciones por COVID-19 en lugares que podría pensarse que los fallecimientos presentarían comportamientos muy distintos, como China y Korea del Sur por un lado, o lugares como Italia, por el otro. En Korea del Sur y China, el control es muy estricto y la cuantificación de casos ha sido masiva. En cambio en Italia, donde la estructura poblacional es distinta, ha habido más fallecimientos por COVID-19, y ha sido rebasado el sistema de salud al grado de tener que negar el uso de respiradores a la gente. Sin embargo, en los tres países el cociente de fatalidad de casos (CFR por sus siglas en inglés) tiene muchas similitudes (Fig. 2, izquierda arriba), que se pueden explotar para obtener las contribuciones relativas de los fallecimientos

por COVID-19 (Fig. 2, izquierda abajo) por cada grupo de edad al total de muertes observadas, y usar esas estimaciones.

Por su similitud, es posible usar los pesos relativos de las defunciones de los distintos grupos de edad y hacer distintas predicciones para el caso de México (Fig. 2).

Las proyecciones obtenidas usando los datos de los distintos países son similares. Hay que tomar en cuenta que estos datos no han sido ajustados con respecto a subreporte. Por ejemplo, ajustando los datos con un factor de 12 por subreporte, la estimación de muertes de adultos de más de 70 años para México es de alrededor 4000 para el 19 de abril de 2020.

Es importante mencionar que estas estimaciones no toman en cuenta la estructura poblacional en México, o la de los países tomados para el análisis, pero esa información está normalizada en el cálculo de los cocientes de fatalidad por caso (Fig. 2).



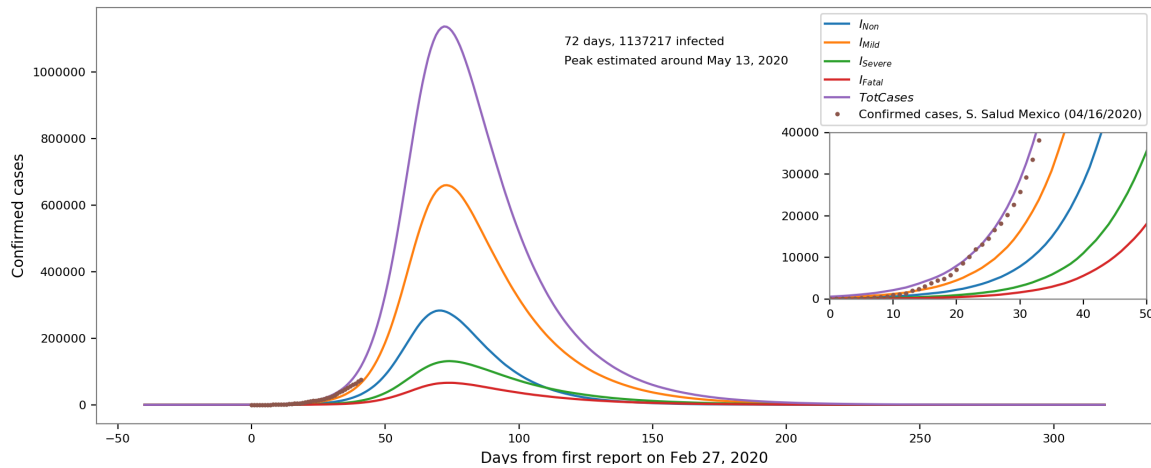
**Figura 3:** Periodos infecciosos para distintos tipos de casos de COVID-19.

## Estimación de fecha de inicio y pico de la epidemia de COVID-19 en México

Por el momento, hemos encontrado regímenes de parámetros al rededor de los cuales podemos simular la dinámica macroscópica de la epidemia en México ajustada con los datos de la Secretaría de Salud de México (Fig. 4). Por construcción, sólo una proporción de los susceptibles y una proporción de los infectados están en contacto. Con el fin de dar estimaciones conservadoras de inicio, hemos realizado simulaciones en las que la

# Covid-19 dynamics from the contribution of infectious individuals with different contagion intervals

(Herrera-Nolasco, Herrera-Valdez, 2020)



**Figura 4:** Estimación de la fecha de inicio y pico de la epidemia ajustando los datos iniciales de la epidemia de Covid-19 en México. La fecha cero representa el 27 de febrero de 2020. Para estas simulaciones los infectados de los grupos severo y crítico en las simulaciones tienen exposición nula a otros grupos poblacionales.

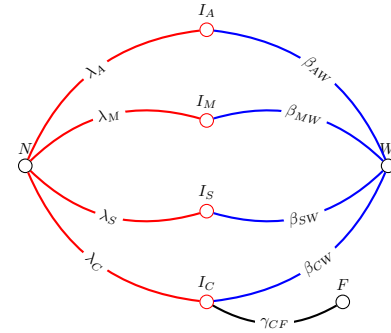
exposición de susceptibles al contagio es constante, ajustando los datos de la Secretaría de salud con un factor de 12, y suponiendo una población expuesta al 10 %. En esas condiciones, una estimación *conservadora* de la fecha de inicio de la epidemia es aproximadamente 40 días antes del 27 de febrero. Es decir, aproximadamente el 20 de enero de 2020. El pico de la infección estaría estimado para ocurrir al rededor de 75 días después del día 0. Es decir, al rededor de Mayo 14 de 2020. La tendencia de los datos de los últimos días es de crecimiento menor en comparación a las simulaciones. Esto indica que posiblemente haya un retraso aún mayor para observar el pico de la epidemia y con un número de casos menor.

En esta dirección, ha resultado útil un caso particular del modelo en las ecuaciones (1)-(3) para explorar escenarios en grandes números. Para ello, tomamos los valores esperados de los muestreos del modelo estocástico y derivamos una ecuación determinista para el régimen en el que los tamaños poblacionales son grandes (Fig. 4). Como resultado obtenemos un modelo que tiene similitud con un SIR clásico.

De forma explícita, si el tamaño de la población  $T = N + I + W$  considerada en el sistema (1)-(3) es lo suficientemente grande, es posible derivar un sistema de ecuaciones diferenciales ordinarias para las densidades poblacionales de no-infectados, infectados y removidos, representadas por  $x$ ,  $\vec{y}$ , y  $w$ , de tal forma que

$$\begin{aligned}\partial_t x &= -\lambda x \\ \partial_t y &= \lambda x - \vec{\gamma} \cdot \vec{y} - \delta_F y_C \\ \partial_t w &= \vec{\gamma} \cdot \vec{y} + \delta_F y_C\end{aligned}$$

donde  $\vec{y}$  y  $\vec{\gamma}$  son vectores que representan, respectivamente, los tamaños de la subpoblaciones infecciosas y las tasas de remoción para cada grupo infeccioso. La variable  $y$  representa la prevalencia (diaria) de infecciones en la población en el tiempo  $t$ .



**Modelo estocástico para definición de intervalos y rangos de acción** (trabajo en progreso) Sea

$$N(t+h) = N(t) - X(t) \quad (1)$$

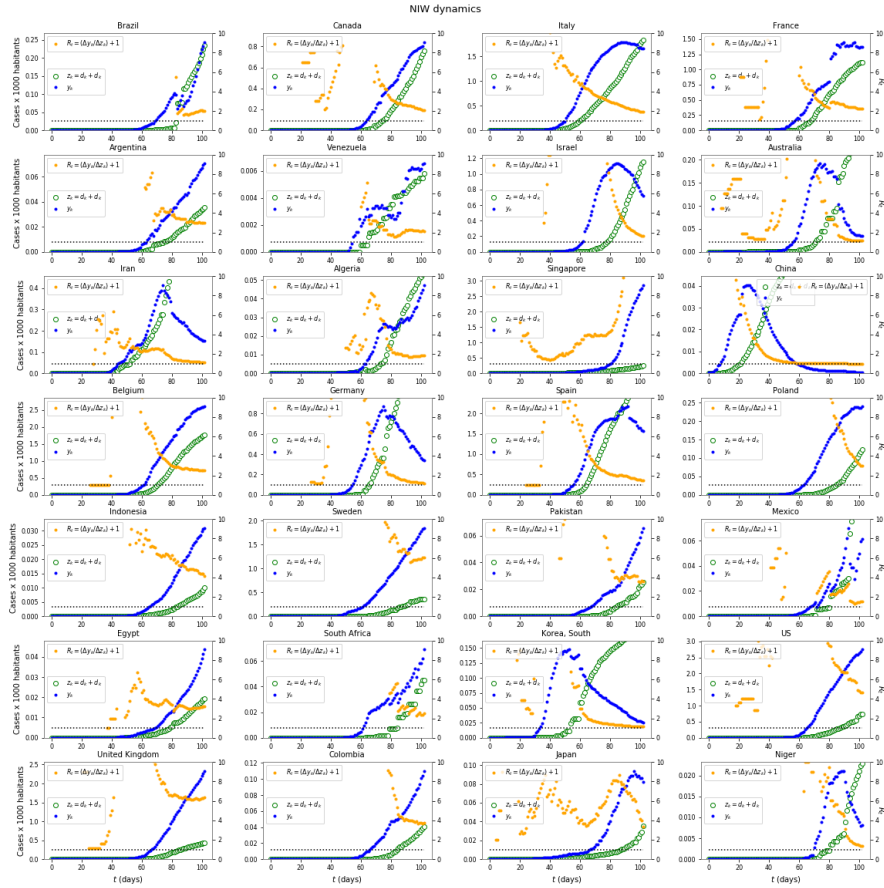
$$I(t+h) = X(t) - Y(t) \quad (2)$$

$$W(t+h) = W(t) + Y(t) - D(t) \quad (3)$$

donde  $X(t)$ ,  $Y(t)$ , y  $D(t)$  son variables aleatorias entera que representan, respectivamente, la incidencia de infección, los nuevos casos removidos de la cadena de infección al cumplir su periodo infeccioso (recuperados), y los nuevos fallecidos. Los miembros del grupo  $W$  ya no participan en la cadena de transmisión. El total poblacional  $N + I + T$  no es constante. La dinámica de la epidemia está supuesta como un proceso estocástico no homogéneo en el que no necesariamente hay mezcla homogénea entre infectados y no infectados (Herrera-Nolasco, 2020, En revisión). Las simulaciones están basadas en muestreos binomiales o Poisson, dependiendo de la cantidad de ensayos requeridos en cada muestreo.

La probabilidad de éxito para cada muestreo puede depender de la cantidad de infectados y su posible contribución (por exposición) a la cadena de transmisión (para  $X$ ), o de tiempos de recuperación y posiblemente muerte (para  $Y$ ). El modelo mostrado en las ecuaciones (1)-(3) también permite simular dinámicas epidémicas en poblaciones pequeñas, y controlar de forma independiente factores como la exposición de los susceptibles al contagio y la contribución de los infectados de distintos grupos, lo que permite considerar escenarios de distanciamiento social, aislamiento de casos sospechosos, aislamiento de casos y la posible contribución de grupos de individuos asintomáticos.

Es importante recalcar que los periodos de infectividad de los casos de COVID-19 son de un mínimo de 2 semanas para el caso asintomático, y hasta 25 días para casos graves (Liu et al., 2020; Zhou et al., 2020). Por tanto, las tasas de remoción del grupo de los infectados deber de tomar en cuenta periodos por lo menos comparables.

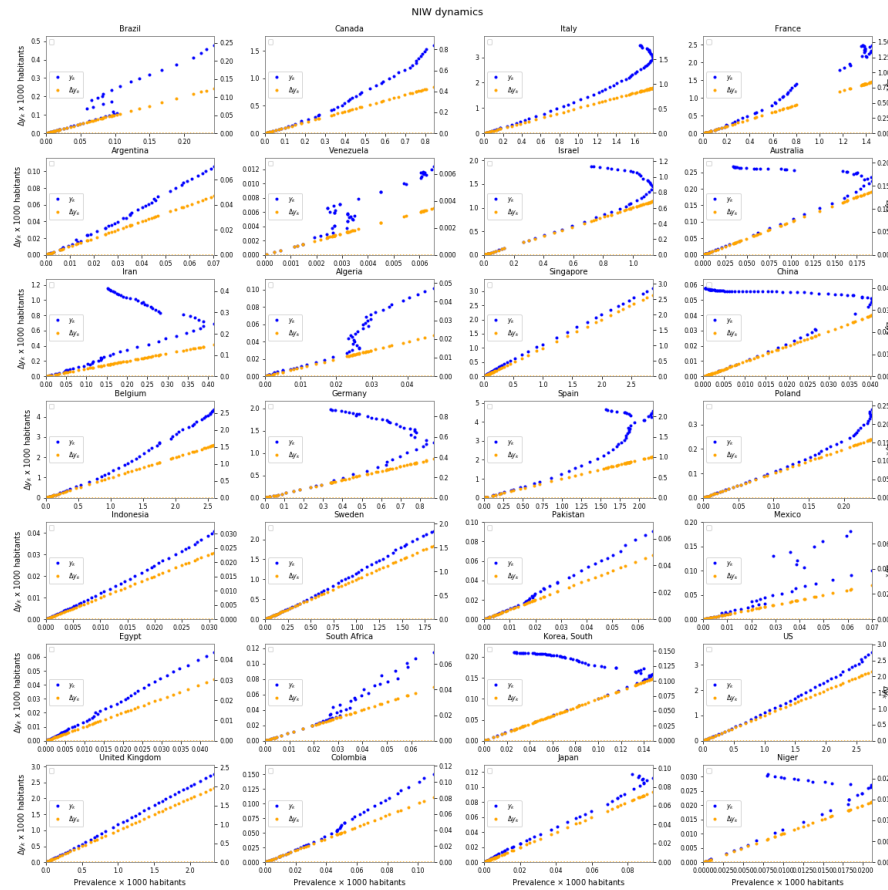


**Figura 5:** Prevalencia (puntos azules), remoción (círculos verdes), y número efectivo de reproducción (puntos naranjas) de COVID-19 en distintos países del mundo. Países como México, Venezuela, Algeria, y Sudáfrica resaltan por su comportamiento cualitativamente distinto al resto de los países mostrados.

## Monitoreo de casos, recuperaciones y defunciones desde una perspectiva teórica, global y en México

Después de un análisis preliminar sobre retrasos entre reportes y contribuciones a las defunciones de grupos de edad, es notorio que comportamiento de la prevalencia de casos en distintos lugares del mundo es muy parecido (Fig. 5), pese a diferencias en sistemas de salud, estructuras poblacionales, métodos de intervención y vigilancia epidemiológica, etcétera. En particular, las trayectorias de las curvas de casos confirmados, recuperados, y defunciones muestran comportamientos similares a los predichos desde una perspectiva cualitativa con modelos simples de tipo SIR (Fig. 6). Lo anterior es cierto para varios países que llevan ya varias semanas reportando casos, y constituye una evidencia de que la dinámica macroscópica (a

nivel de país) no varía mucho en el mundo, independientemente de las particularidades de cada lugar.



**Figura 6:** Retrato de fase entre prevalencia e incidencia (naranja) y entre prevalencia y cambio en la prevalencia (azul), en distintos países del mundo. Compare los retratos de fase de México, Venezuela, Algeria, y Sudáfrica con el resto de la muestra.

Primero caracterizamos la dinámica epidémica en términos generales, tomando en cuenta datos globales, para después hacer estimaciones de carácter puntual que permitan entender mejor el caso de México y plantear estrategias de mitigación, además de estimaciones de eventos importantes como máximos locales de prevalencia, inicio de la epidemia, etcétera. Para hacerlo, tomamos datos de distintos lugares del mundo, estudiamos su dinámica desde una perspectiva teórica (en consideración de predicciones cualitativas hechas por modelos), y obtenemos patrones de comportamiento que nos permitan modelar y estimar distintos escenarios. El último punto se basará en modelaje de la dinámica de COVID-19 a nivel de país, y luego para el caso de México, tomando en cuenta regiones en todo el país.

**¿Es la dinámica de propagación de COVID-19 observada en distintos países similar a la predicha en modelos?** A partir de datos globales ([https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\\_covid\\_19\\_data/csse\\_covid\\_19\\_time\\_series/](https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/)), calculamos el número efectivo de reproducción, las prevalencias, incidencias de los datos de distintos países<sup>5, 6</sup> y), incluyendo México.

La dinámica macroscópica de propagación en varios países tiene las características esenciales que predicen el modelo NIW (non-infected, infected, withdrawn por sus siglas en inglés) como el propuesto aquí. En aquellos países donde no es así, cabe preguntarse si los reportes, y no la epidemia, están causando que la tendencia predicha por un modelo simple predice.

En breve, si la population se divide en tres subconjuntos representando no infectados y susceptibles, los infectados, y los que dejaron de ser infectados porque se recuperaron y adquirieron inmunidad, o murieron (ya no participan en la cadena de infección). Supongamos que las densidades de los tres grupos son respectivamente  $x$ ,  $y$ , y  $w$ , con  $1 = x + y + w$  y que  $\alpha$  y  $\tau$  representan la tasa de infección dado un contacto infeccioso, y el tiempo promedio de espera para dejar de estar en el grupo infeccioso. La dinámica de  $(x, y, w)$  se puede describir usando el SIR clásico como

$$\partial_t y = \alpha y \left( 1 - w - y \right) - \frac{y}{\tau}, \quad (4)$$

and the evolution for  $z$  is then

$$\partial_t w = \frac{y}{\tau} \quad (5)$$

con  $\partial_t x = -\partial_t y - \partial_t w$ . La incidencia de casos (los casos nuevos) es entonces

$$\partial_t y + \partial_t w = \alpha y \left( 1 - w - y \right) \quad (6)$$

Entonces la incidencia es una función cuadrática de la prevalencia, pero también depende de los removidos de la cadena de infección (muertos e inmunes). Como  $w$  es una función creciente, la gráfica de la incidencia se puede pensar como una curva en 3D con forma cuadrática. En teoría, se puede entonces calcular

$$\alpha = \frac{\partial_t y + \partial_t w}{\partial_t w} \quad (7)$$

y

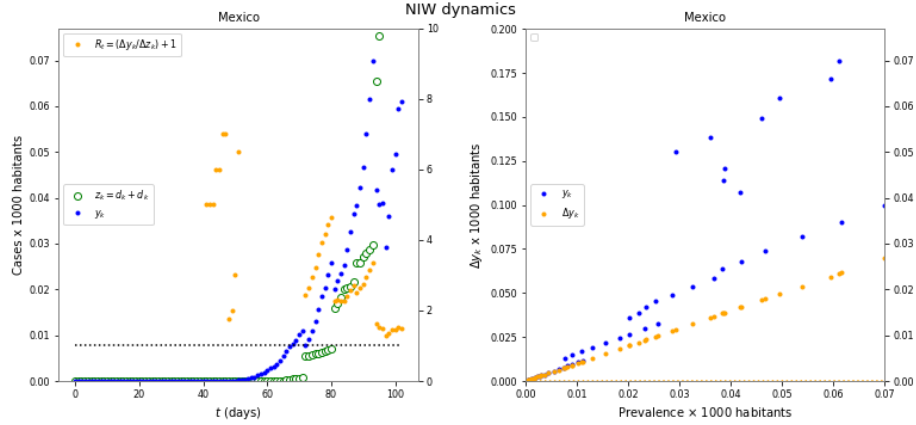
$$\tau = \frac{y}{\partial_t w}. \quad (8)$$

*Parámetros para modelar a partir de datos.*

De forma parecida a la técnica propuesta por [Lega and Brown \(2016\)](#), y con base en el desarrollo explicado arriba, es posible obtener series de tiempo  $y_k : k = 1, \dots, d$  para la prevalencia,  $z_k : k = 1, \dots, d$  para la incidencia, y  $w_k : k = 1, \dots, d$  para los removidos, con  $d$  representando los días desde el inicio de la pandemia.



**Figura 7:** Dinámica de epidemia de COVID-19 en México. Dinámica de prevalencia (puntos azules), remoción (muertes y recuperaciones de casos confirmados, círculos verdes), y número efectivo de reproducción (puntos naranjas). Retrato de fase entre prevalencia e incidencia (naranja) y entre prevalencia y cambio en la prevalencia (azul).



Con esas series de tiempo, se pueden calcular coeficientes

$$a_k = \frac{z}{\delta w}, \quad (9)$$

y

$$b(t) = \frac{y}{\delta w}. \quad (10)$$

análogos a  $\alpha$  y  $\tau$  con los que puede construirse un modelo con coeficientes variables en el tiempo. En particular, se puede *estimar* el número efectivo de reproducción dividiendo la incidencia entre el cambio en los removidos (Fig. 7). **El número efectivo de reproducción en México en la última semana de abril y la primera de mayo está entre 1.64 y 1.5.**

Para fines de comparación, considere las dinámicas de Corea del Sur, China (Fig. 8) y en Italia y España (Fig. 9).

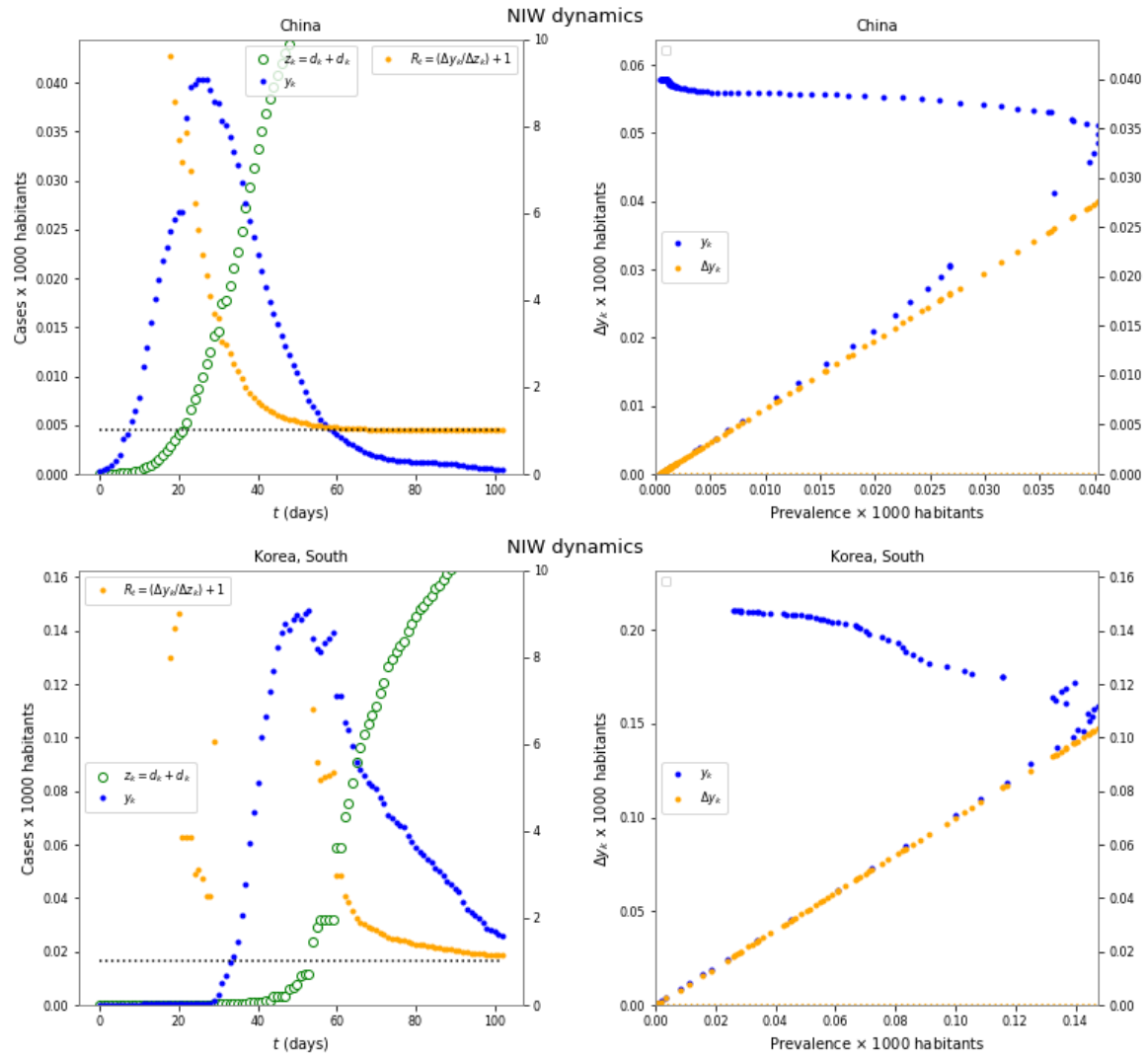
**Trabajo en progreso.** La inferencia sobre características globales de la epidemia y propiedades específicas sobre el COVID-19 a nivel de un sólo país, inclusive desde perspectivas de provincias (estados en el caso de México) está siendo abordada por nuestro equipo con la idea de calibrar modelos para explicar la dependencia de la propagación con respecto a grupos de distintas vulnerabilidades, y también para proponer y evaluar estrategias de mitigación.

## Referencias

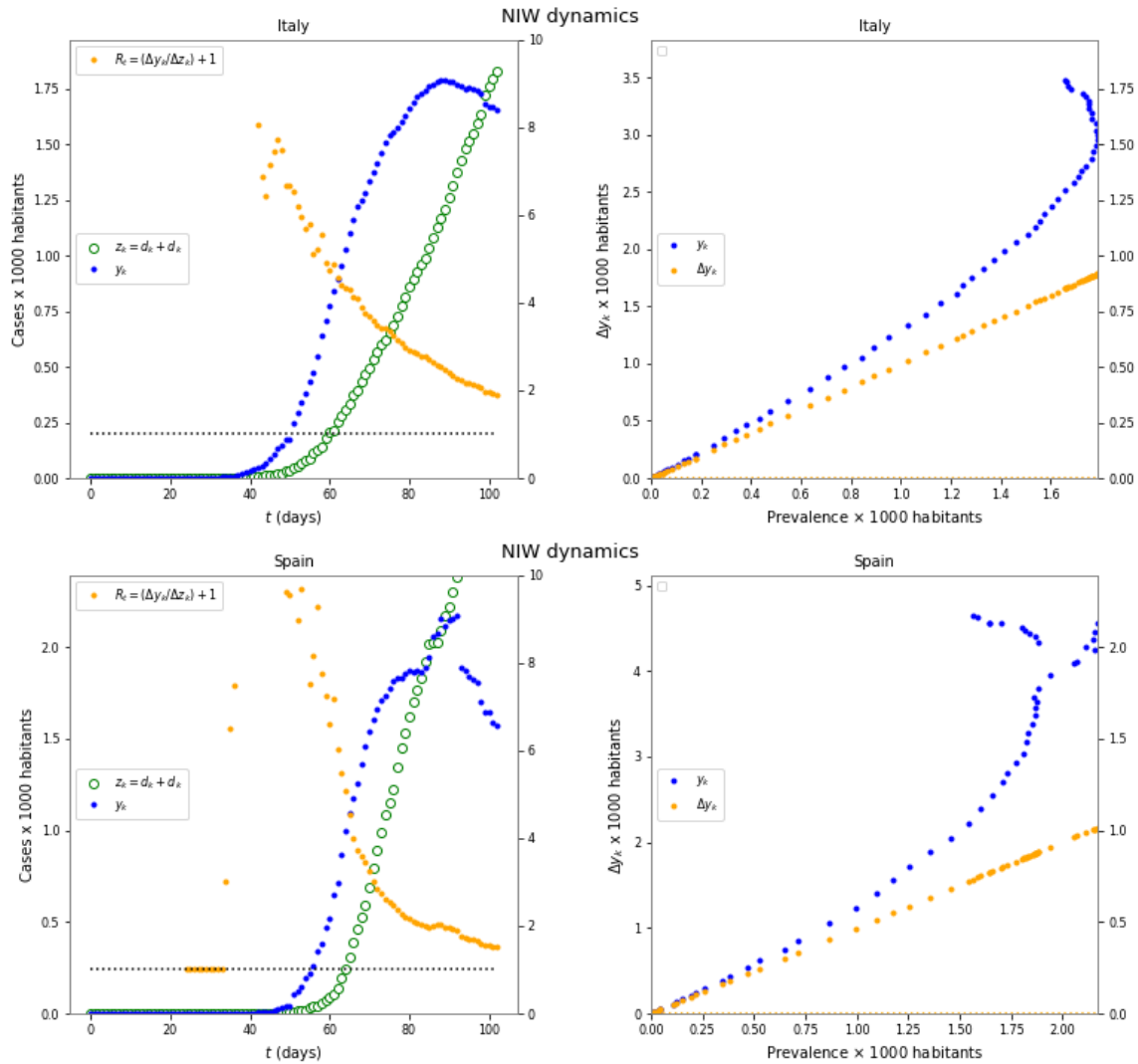
Joceline Lega and Heidi E Brown. Data-driven outbreak forecasting with a simple nonlinear growth model. *Epidemics*, 17:19–26, 2016.

Yang Liu, Li-Meng Yan, Lagen Wan, Tian-Xin Xiang, Aiping Le, Jia-Ming Liu, Malik Peiris, Leo LM Poon, and Wei Zhang. Viral dynamics in mild and severe cases of covid-19. *The Lancet Infectious Diseases*, 2020.

Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, et al. Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *The Lancet*, 2020.



**Figura 8:** Epidemia de COVID-19 en China y Corea del Sur. Dinámica de prevalencia (puntos azules), remoción (muertes y recuperaciones de casos confirmados, círculos verdes), y número efectivo de reproducción (puntos naranjas). Retrato de fase entre prevalencia e incidencia (naranja) y entre prevalencia y cambio en la prevalencia (azul).



**Figura 9:** Epidemia de COVID-19 en Italia y España. Dinámica de prevalencia (puntos azules), remoción (muertes y recuperaciones de casos confirmados, círculos verdes), y número efectivo de reproducción (puntos naranjas). Retrato de fase entre prevalencia e incidencia (naranja) y entre prevalencia y cambio en la prevalencia (azul).