

Análisis de la evolución del COVID-19 mediante filtros de Kalman

Antonio Gómez Expósito, Catedrático de Ingeniería Eléctrica, Universidad de Sevilla
José A. Rosendo Macías, Catedrático de Ingeniería Eléctrica, Universidad de Sevilla
Miguel A. González Cagigal, Becario FPU, Universidad de Sevilla



Nota: Este documento muestra los resultados obtenidos con los datos disponibles a 5 de abril de 2020. Dado que los resultados incluyen otros países, y que se ha añadido nuevo material sobre el modelado matemático, hemos cambiado el título y la versión, que a partir de ahora será “v2”. Se remite al lector interesado a la serie de documentos “v1” publicados desde el 24 de marzo para el caso de España, incluyendo la nota explicativa del 1 de abril, que complementan este documento y que se pueden descargar de: <https://idus.us.es/handle/11441/94508>

Resumen

Puesto que España está muy cerca del pico-meseta de la pandemia, si no lo ha superado ya, el interés principal se centra a partir de ahora en su evolución posterior. En éste y sucesivos documentos intentaremos responder a las siguientes preguntas: ¿Cómo y cuándo volveremos al estado de partida de mantenerse las condiciones actuales? ¿Qué ocurriría si se levantase el confinamiento prematuramente? etc. Además de los resultados, se incluye un nuevo apartado (Anexo) donde se justifican teóricamente las nuevas hipótesis realizadas sobre la evolución futura de $r(n)$.

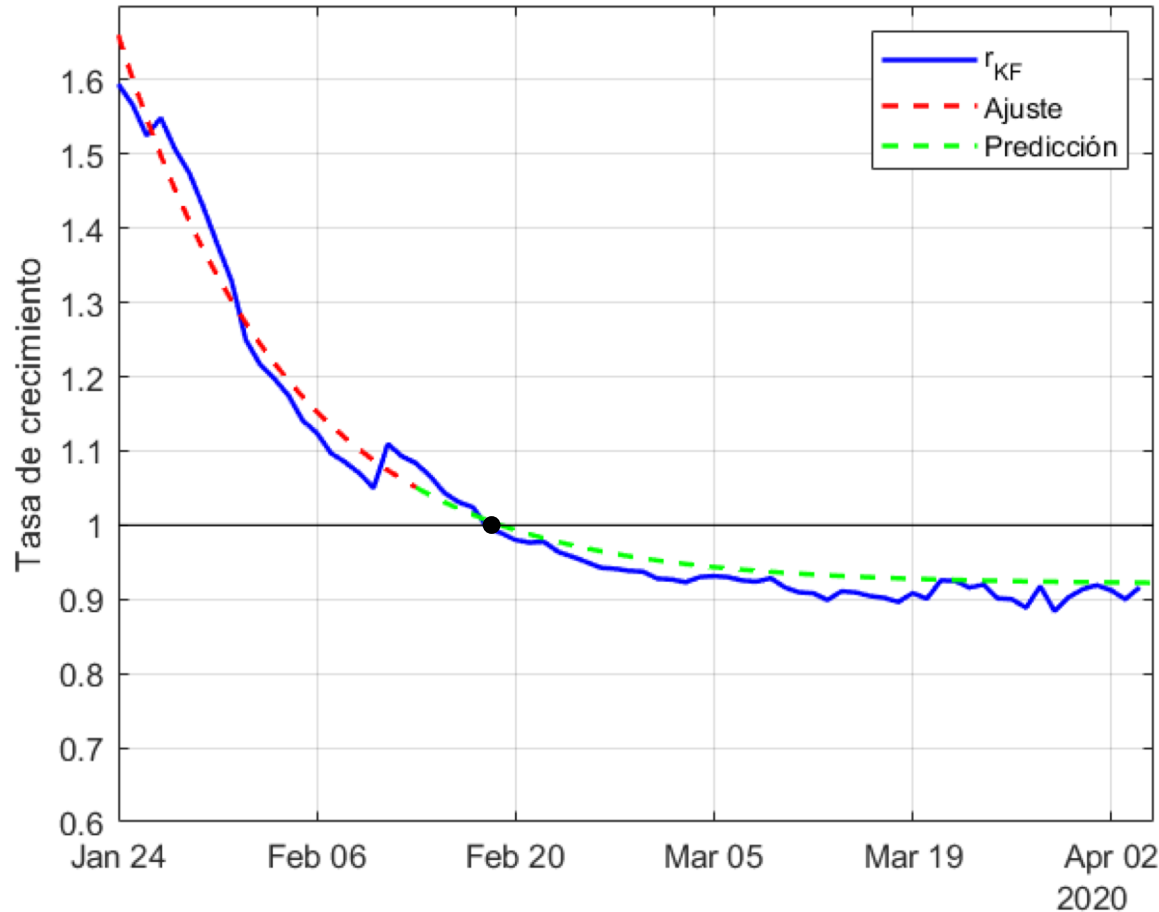
Notación

$r(n)$: razón geométrica variable en el tiempo del número de infectados por el Covid-19 en cada instante de tiempo n . Dicha razón se puede estimar aunque el número de infectados no llegue nunca a saberse con certeza.

$\alpha(n)$: tasa que modela la fracción de infectados que resultan positivos, para considerar la posibilidad real de que haya más infectados que los reportados como positivos.



Validación del método. Caso de China

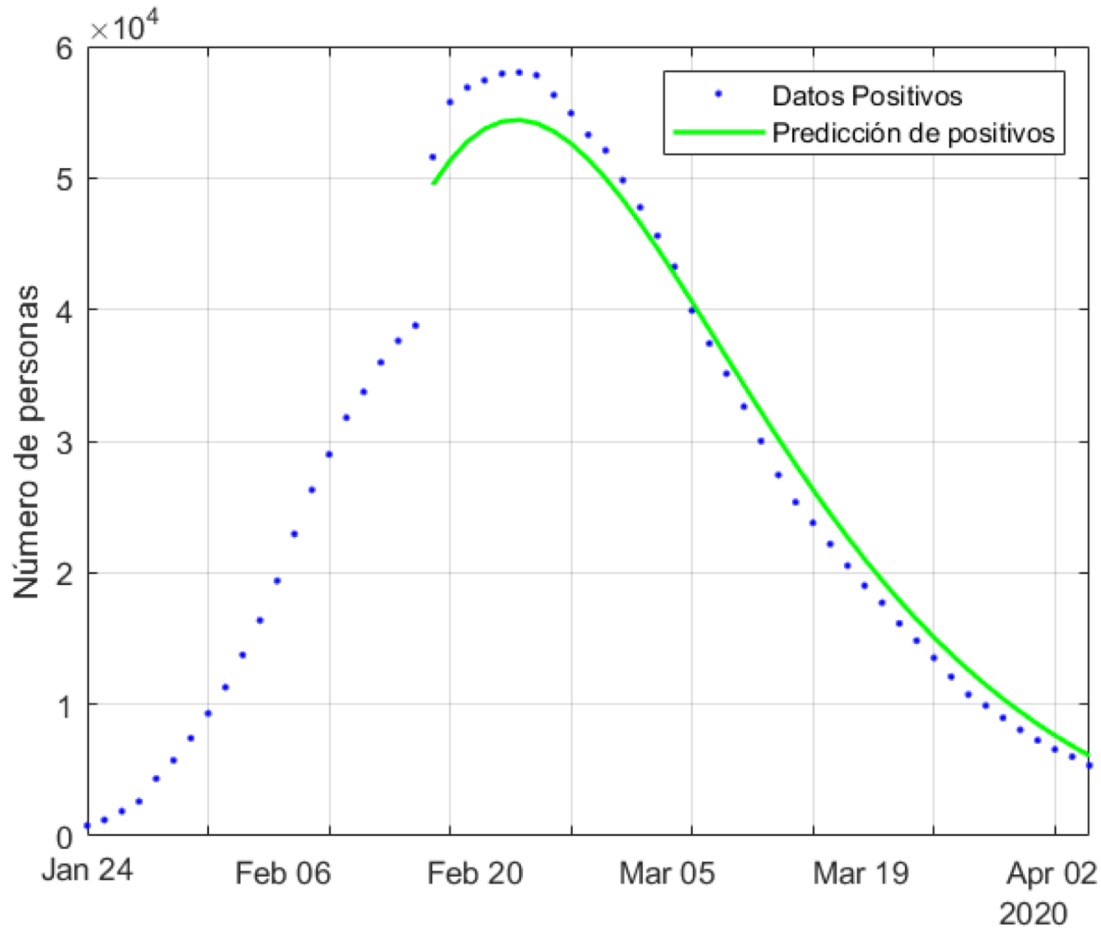


Para contrastar la validez de las hipótesis teóricas desarrolladas en el Anexo, sobre la evolución futura de $r(n)$, se ha considerado el caso de China, la cual sufrió los efectos de la pandemia con anterioridad al resto del mundo, y parece encontrarse actualmente en la fase final de remisión.

Para realizar el ajuste exponencial al que hace referencia el Anexo de este informe, se han tomado los datos existentes hasta 5 días antes del pico (paso por 1 de la tasa de crecimiento), extrapolando dicha curva a los días posteriores (verde).

En la figura, que incluye el valor de $r(n)$ estimado por el filtro de Kalman con los datos reales durante todo el periodo, puede apreciarse que la extrapolación es bastante precisa, prediciéndose bien no sólo el momento del paso por 1 sino la tendencia asintótica a largo plazo.

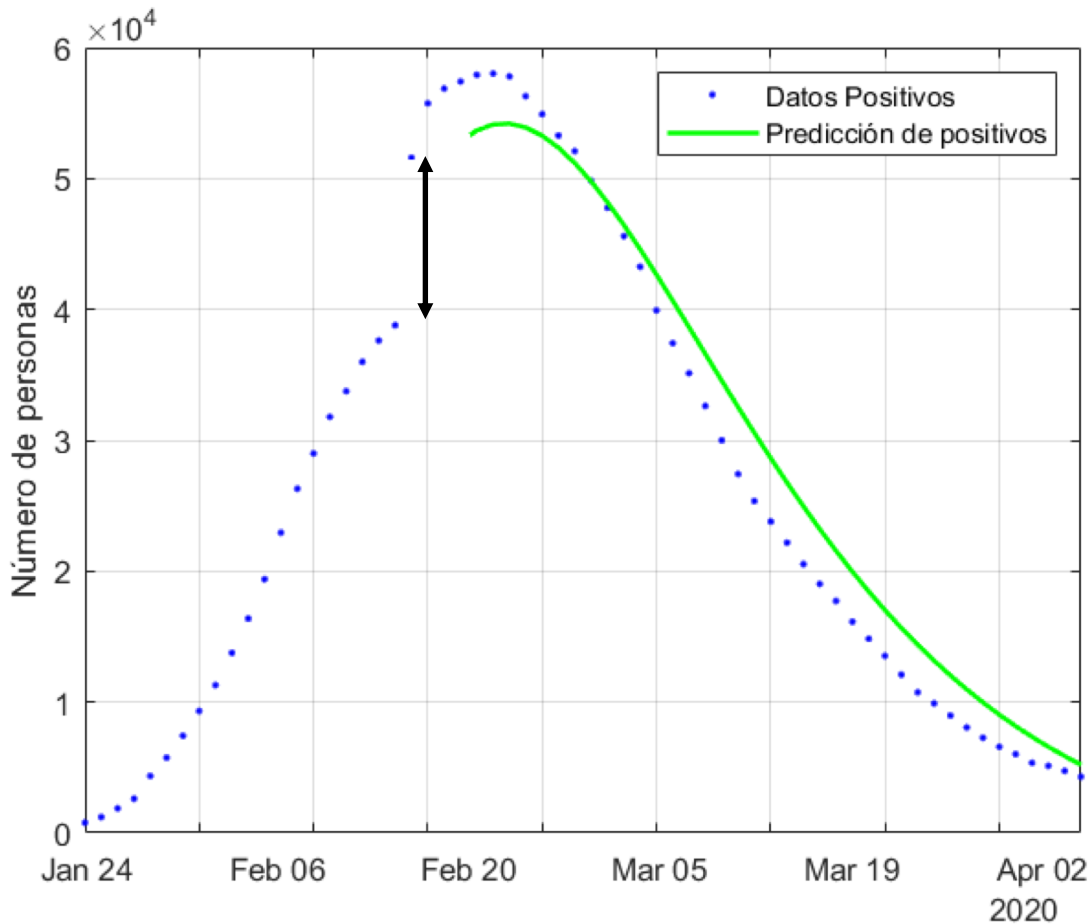
Validación del método. Caso de China



En base a la evolución de la tasa de crecimiento exponencial mostrada anteriormente, se llevó a cabo el cálculo de los positivos activos en China (verde), que se contrastan en esta figura con las cifras reales reportadas en este país.

Puede comprobarse que el método propuesto predice satisfactoriamente la forma de la bajada de la curva de positivos, aspecto que será de vital importancia en los días venideros para la planificación del fin del estado de confinamiento en el resto de países.

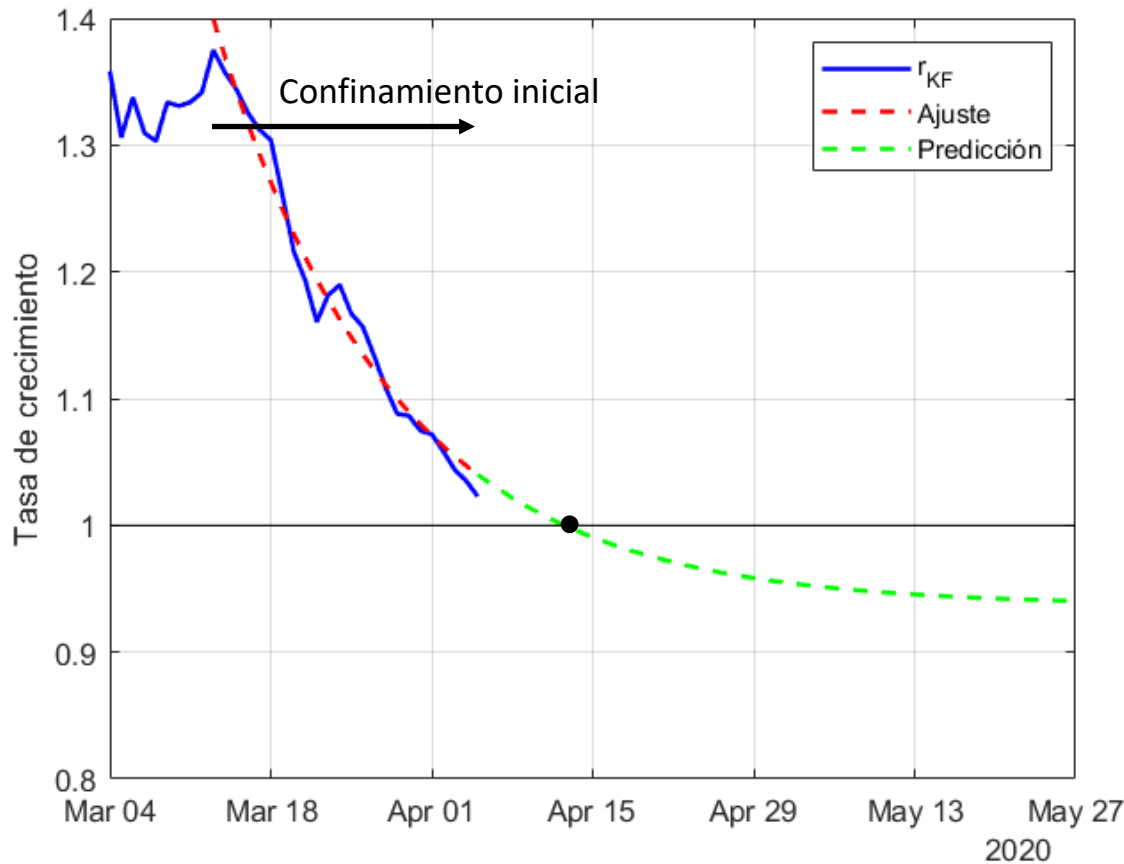
Validación del método. Caso de China



Finalmente, para probar la robustez del método utilizado, se procedió a incluir en el ajuste exponencial de $r(n)$ los datos disponibles hasta 2 días antes del paso por 1 de la tasa de crecimiento (en lugar de 5, como en la figura anterior). De esta forma, se podría notar la influencia de la discontinuidad en el número de positivos reportados que puede apreciarse en China el día 11 de febrero (un incremento de más de 10.000 casos de un día para otro), probablemente ocasionada por un incremento en el número de tests.

Puede comprobarse, a la vista de la figura, que la predicción ha empeorado ligeramente en la bajada. No obstante, esta diferencia no es excesiva y permite validar el método de ajuste y extrapolación, de manera que pueda aplicarse en lo sucesivo al resto de casos considerados.

Análisis de la tasa de crecimiento en España

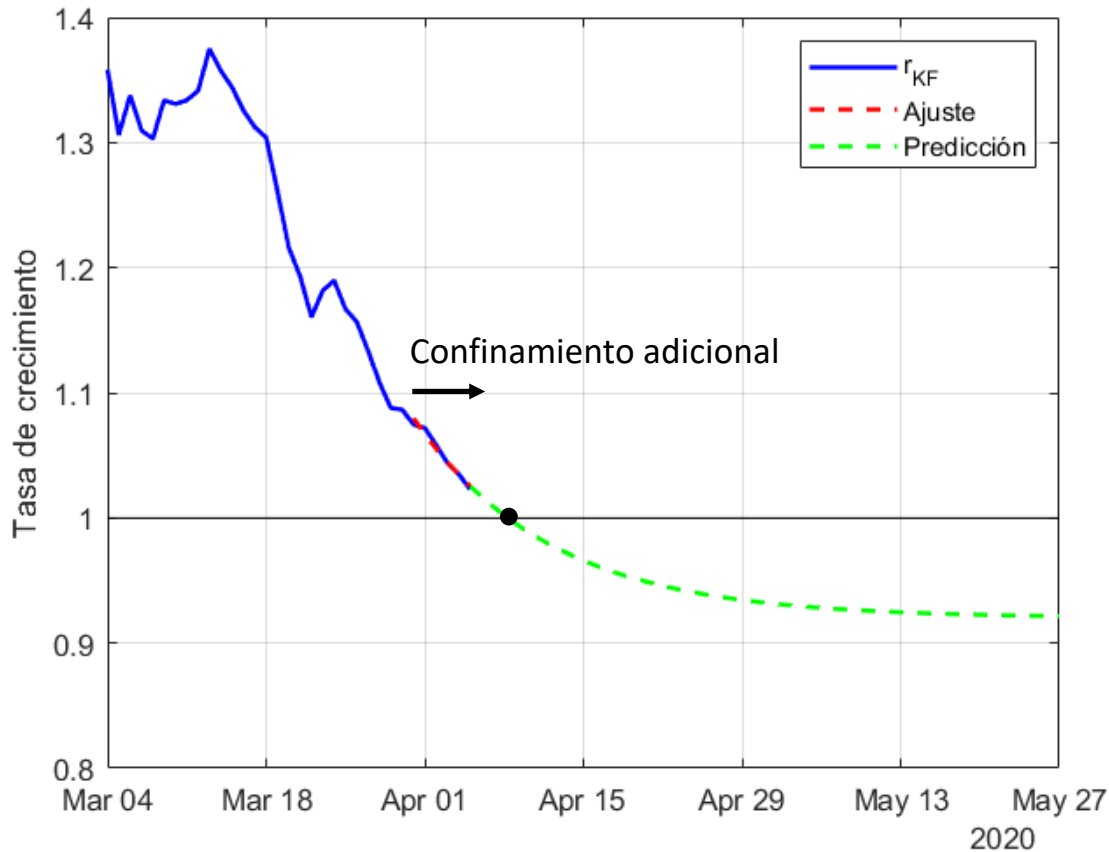


Para la tasa $r(n)$ en el territorio español se ha tomado el mismo ajuste exponencial que el empleado para los datos de China.

La primera ventana de datos utilizada en el cálculo de la curva exponencial que caracteriza la evolución de $r(n)$, incluye los valores reportados desde el día del primer confinamiento en España, el 15 de Marzo, pudiendo verse en la figura que, de acuerdo a dicha extrapolación, el paso por 1 de la tasa de crecimiento (pico) se produciría entre los días 10 y 12 de abril.

Puede verse también, en lo referente a la asíntota de la curva, que su valor es aproximadamente 0,94, superior al de China.

Análisis de la tasa de crecimiento en España

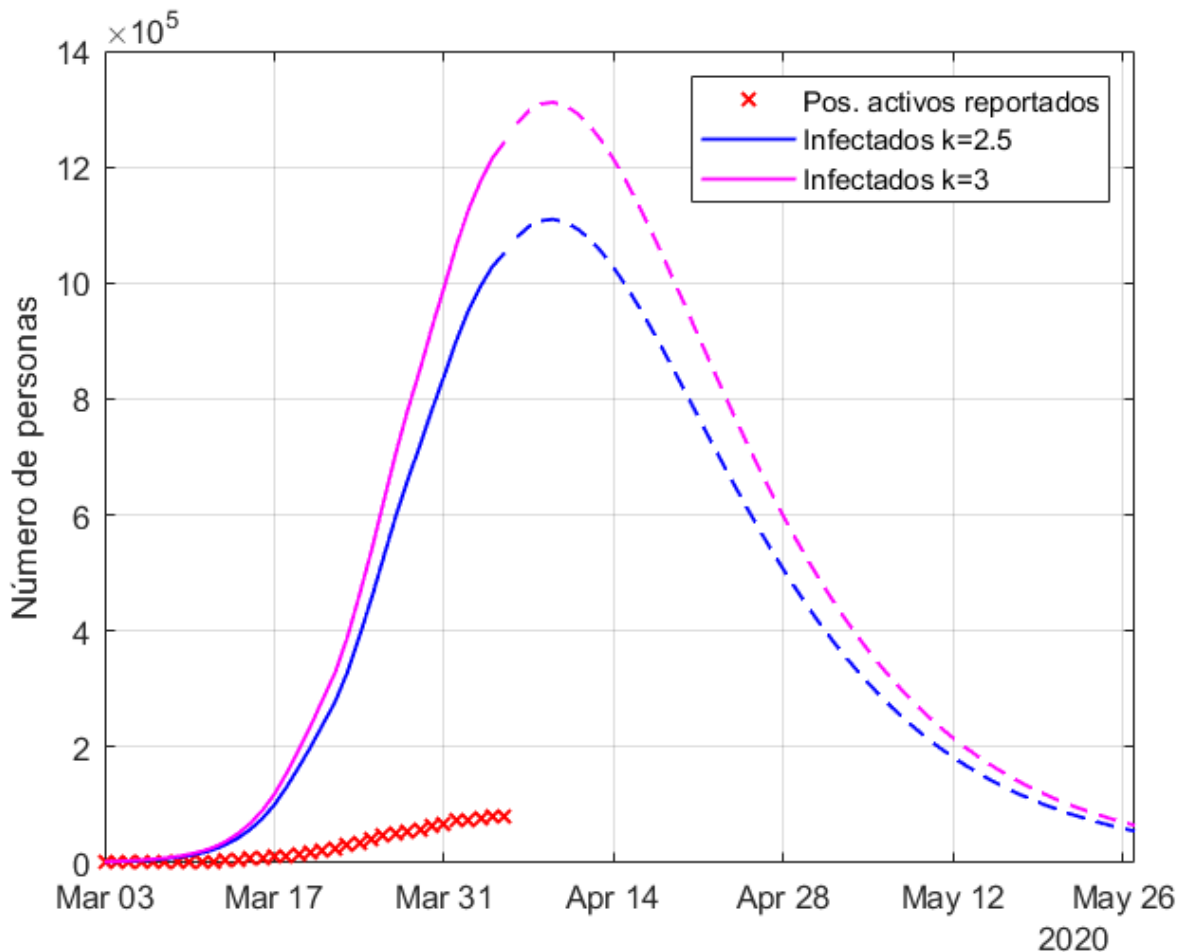


Para analizar la influencia de las nuevas condiciones de confinamiento que se establecieron el pasado día 30 de marzo en la evolución de $r(n)$, se ha recalculado el ajuste exponencial de dicho parámetro, considerando en este caso una ventana correspondiente a los últimos 5 datos disponibles.

El resultado, mostrado en la figura, conduce a una fecha más próxima para el paso por 1 (pico), en torno al día 8 de abril, habiéndose también producido un descenso en el valor de la asíntota, que pasa a estar próxima a 0,92 (similar a la de China).

Puesto que esta ventana de datos refleja de forma más adecuada la realidad actual, hemos considerado pertinente utilizar este ajuste para la obtención de los datos que se van a mostrar a continuación.

Predicción del número de infectados activos en España



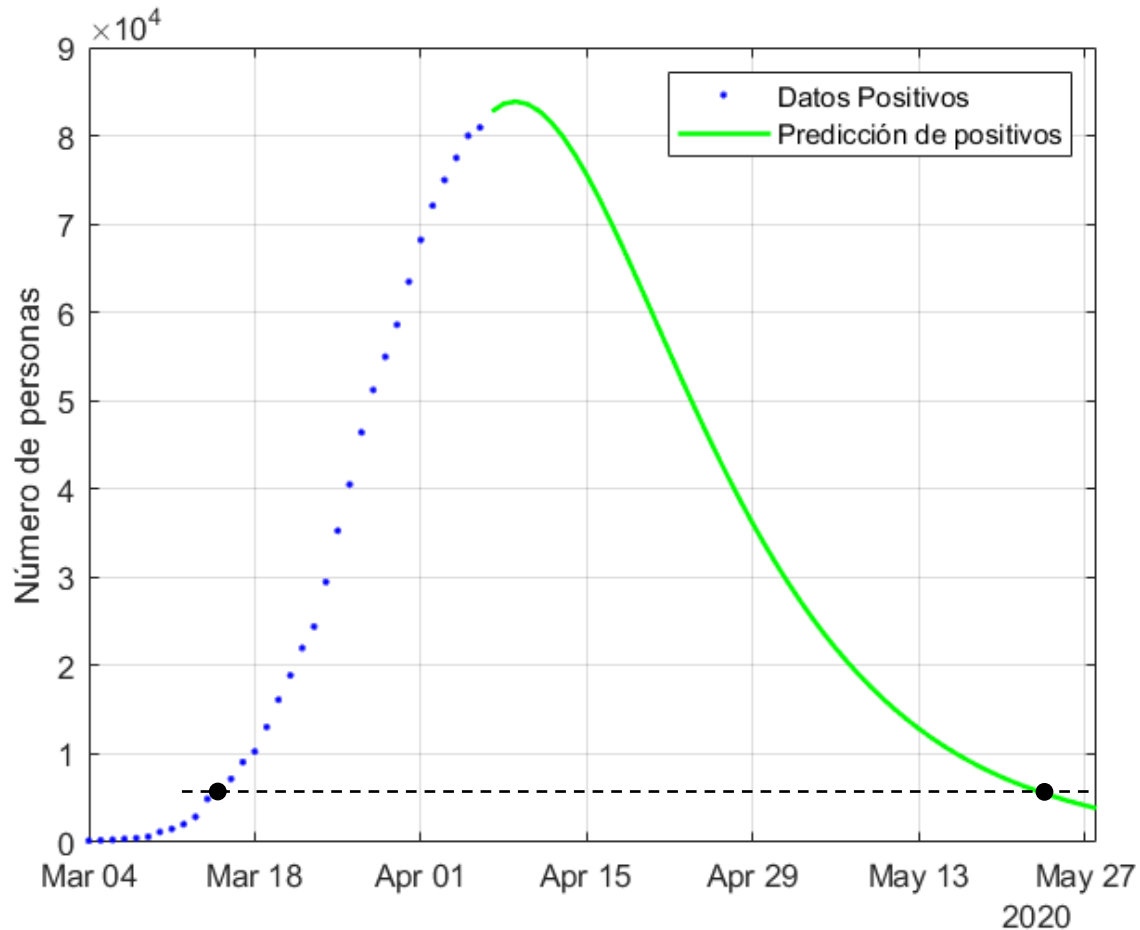
Valores estimados en línea continua y predicciones en discontinua

Con el valor de $r(n)$ estimado y su correspondiente extrapolación a futuro exponencial, se ha calculado el valor estimado y la predicción de los próximos días del número de infectados activos en España, o sea, descontando las personas fallecidas y las recuperadas.

Como en anteriores informes, se ha considerado que el número de infectados en el instante inicial estaba entre 2,5 y 3 veces el número de positivos confirmados en aquel momento (si se supone que dicho número era 5, por ejemplo, bastaría con multiplicar por dos la curva de 2,5).

En el territorio nacional, se prevé que el número máximo de personas infectadas activas esté entre 1,1 y 1,3 millones de personas, para los números iniciales considerados.

Predicción del número de positivos activos reportados en España

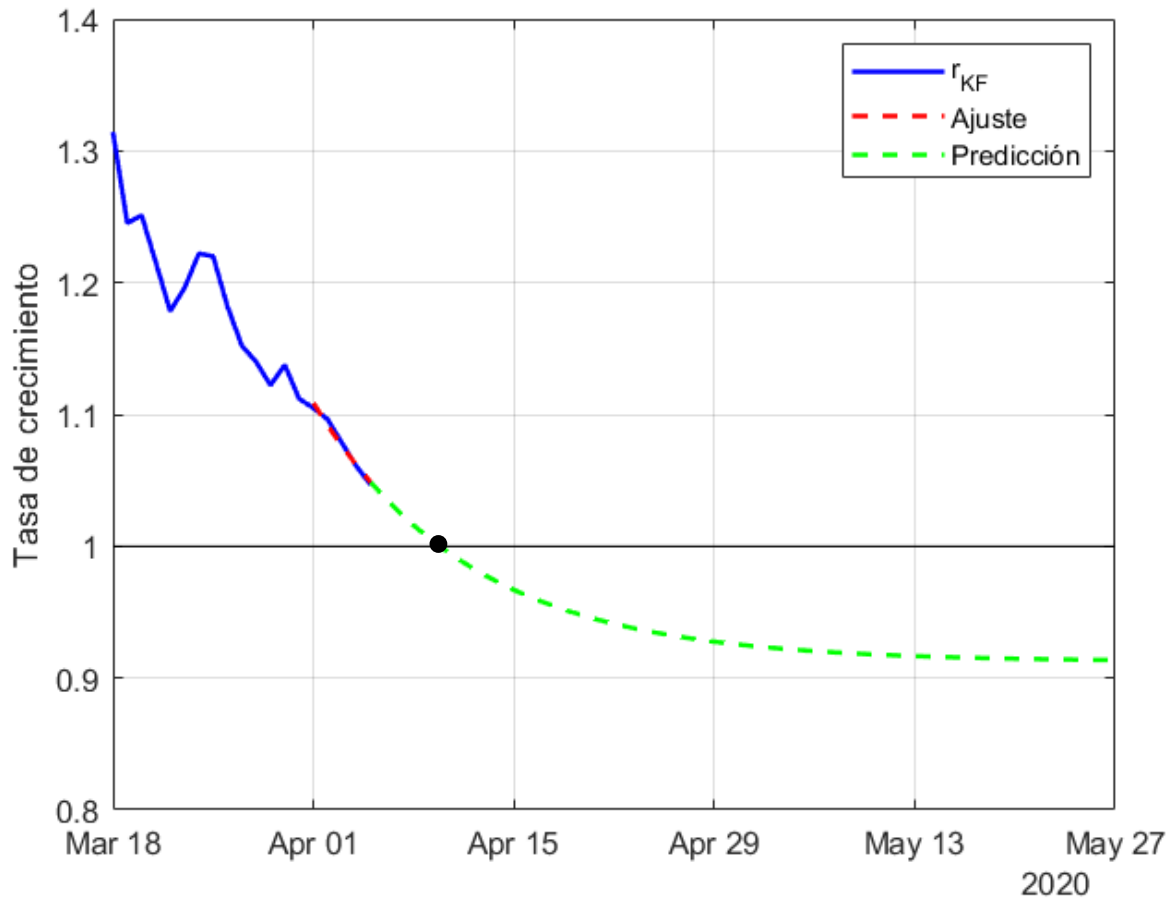


Finalmente, se muestra para el territorio español la predicción del número de positivos activos reportados, pudiendo verse que, en el pico, el número de positivos activos previstos está entre 80 y 90 mil personas.

Un aspecto importante en el estado actual es poder predecir en qué momento el número de personas afectadas será inferior a un determinado valor. En este caso, se ha considerado como referencia el número de positivos reportados el día 15 de marzo, cuando se inició el estado de alarma y primer distanciamiento social.

Este número, correspondiente a 5.678 personas, no se volverá a alcanzar hasta fechas próximas al 23 de mayo (como sugiere la línea discontinua).

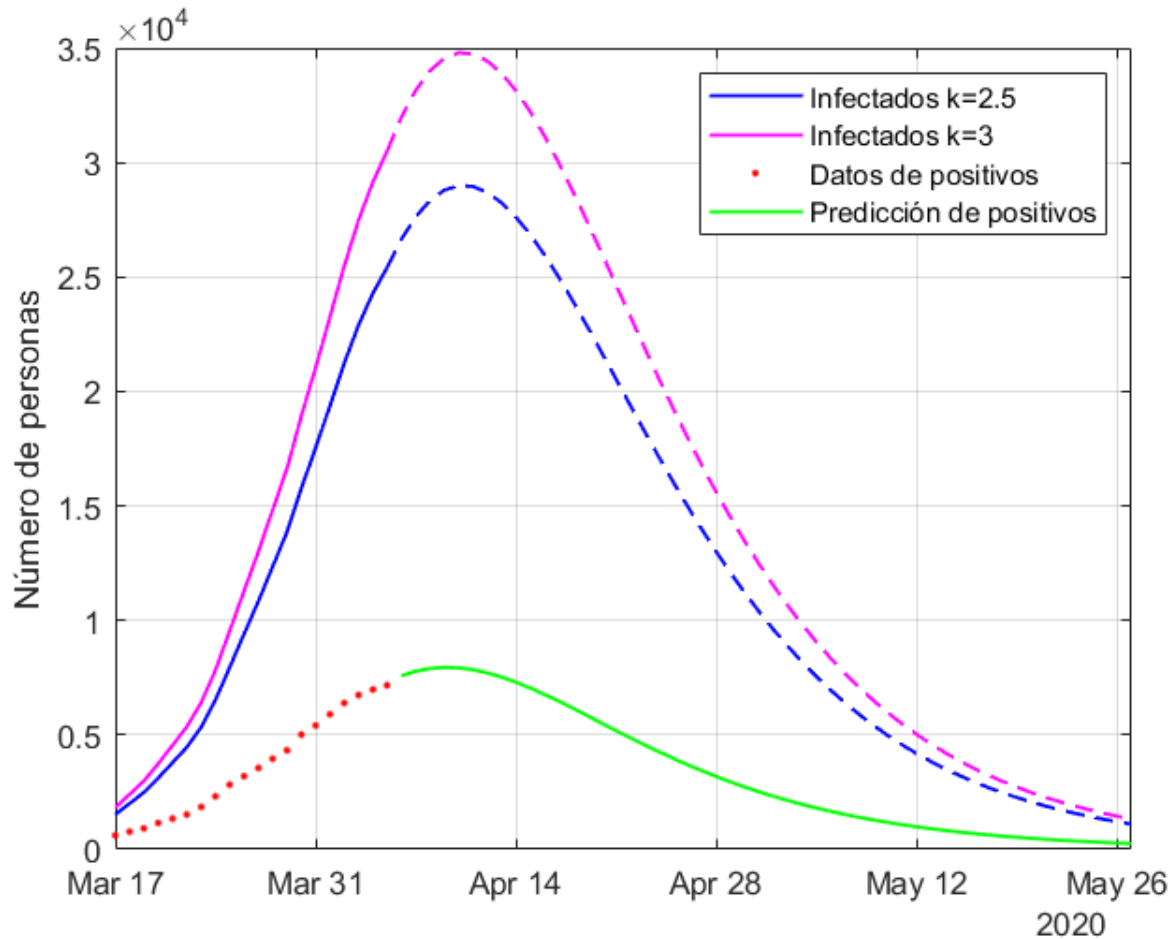
Análisis de la tasa de crecimiento en Andalucía



Un procedimiento análogo al de España se ha empleado para la extrapolación de la tasa $r(n)$ en diversas comunidades autónomas, mostrándose en la figura el caso de Andalucía.

Con una ventana de los últimos 5 datos disponibles, se prevé que el paso por 1 de la tasa de crecimiento se produzca entre el 10 y el 12 de abril.

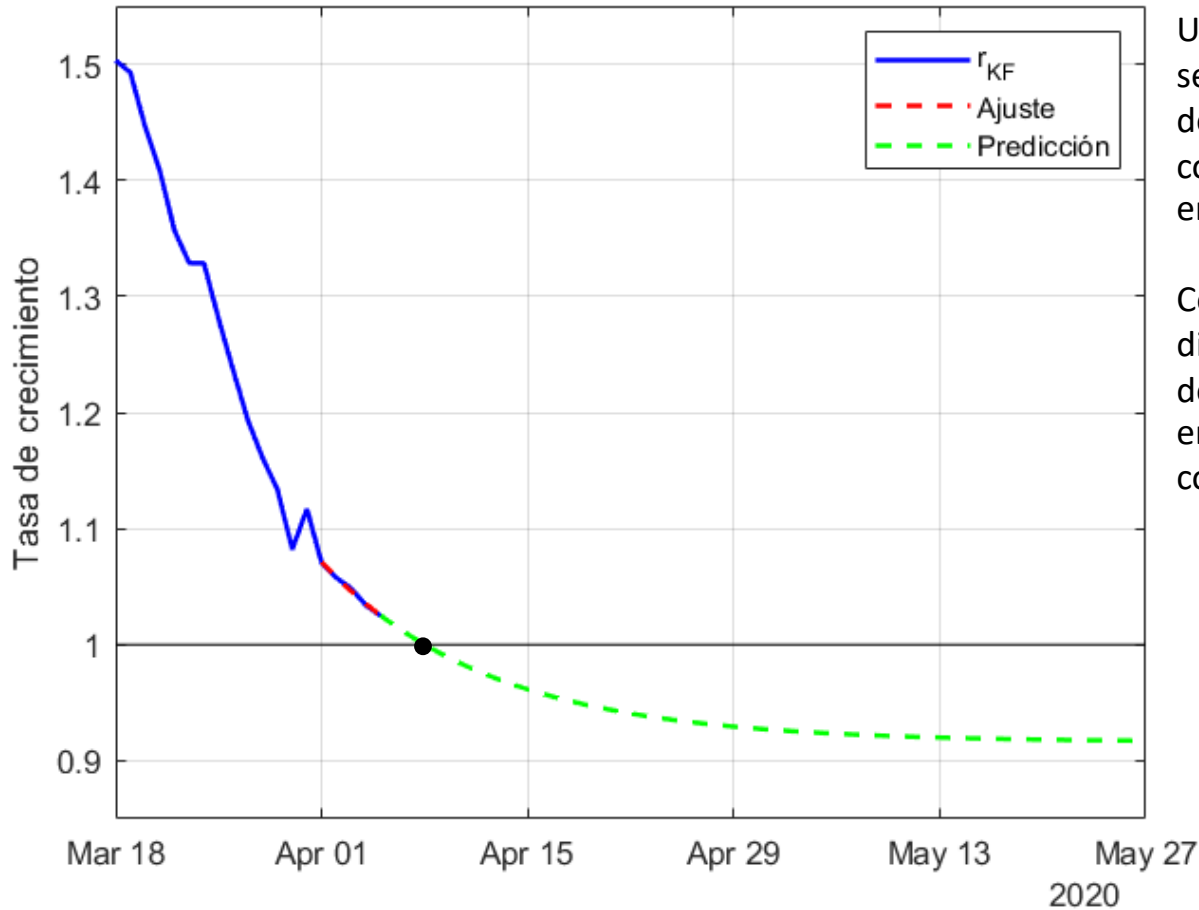
Predicción del número de infectados activos y positivos reportados en Andalucía



Con las mismas consideraciones en lo que concierne al factor k , se muestra en la gráfica adjunta la predicción del número de infectados en la comunidad andaluza.

Conjuntamente a estas curvas se ha añadido también la predicción del número de positivos activos reportados.

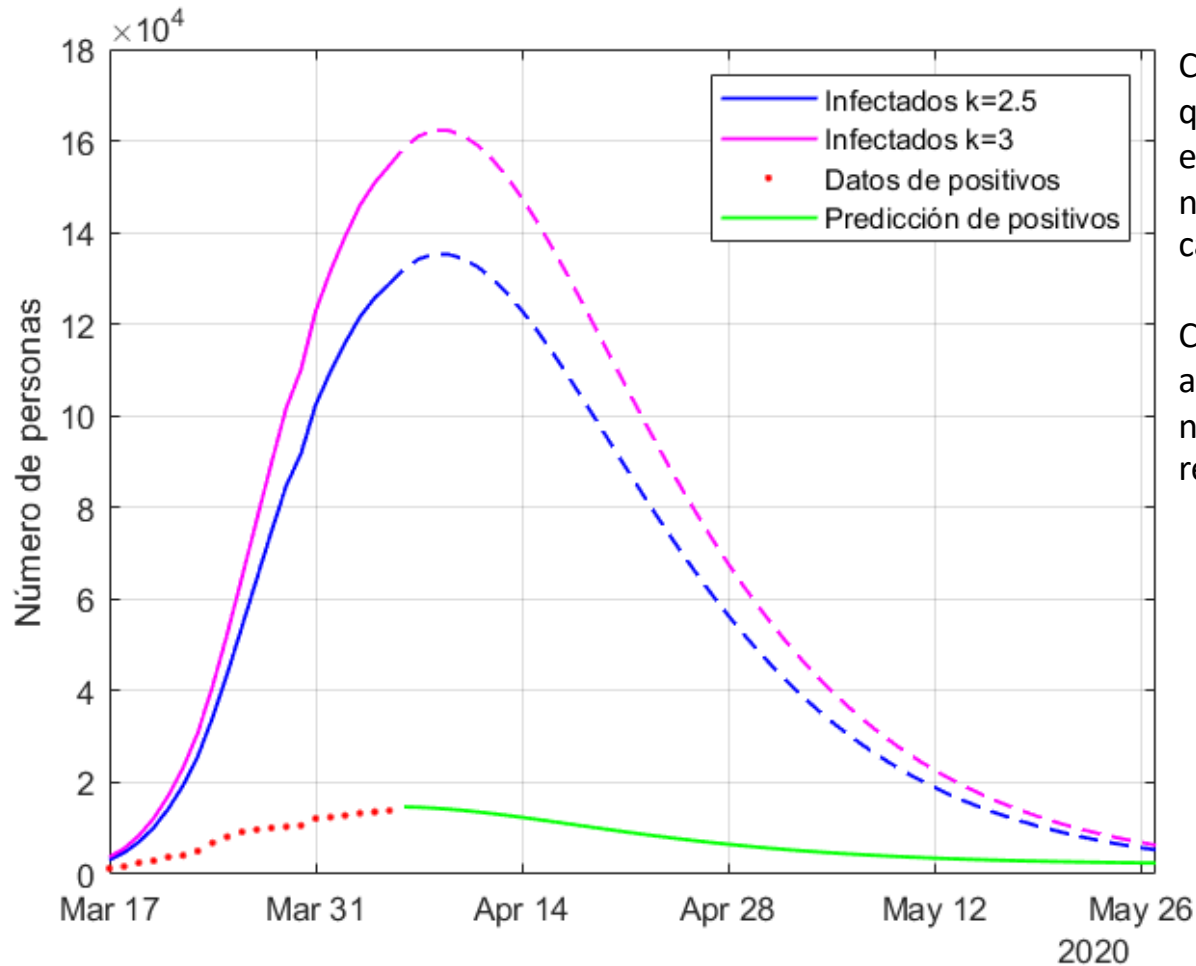
Análisis de la tasa de crecimiento en Cataluña



Un procedimiento análogo al de España se ha empleado para la extrapolación de la tasa $r(n)$ en diversas comunidades autónomas, mostrándose en la figura el caso de Cataluña.

Con una ventana de los últimos 5 datos disponibles, se prevé que el paso por 1 de la tasa de crecimiento se produzca entre el 8 y el 9 de abril, fecha similar al conjunto del territorio nacional.

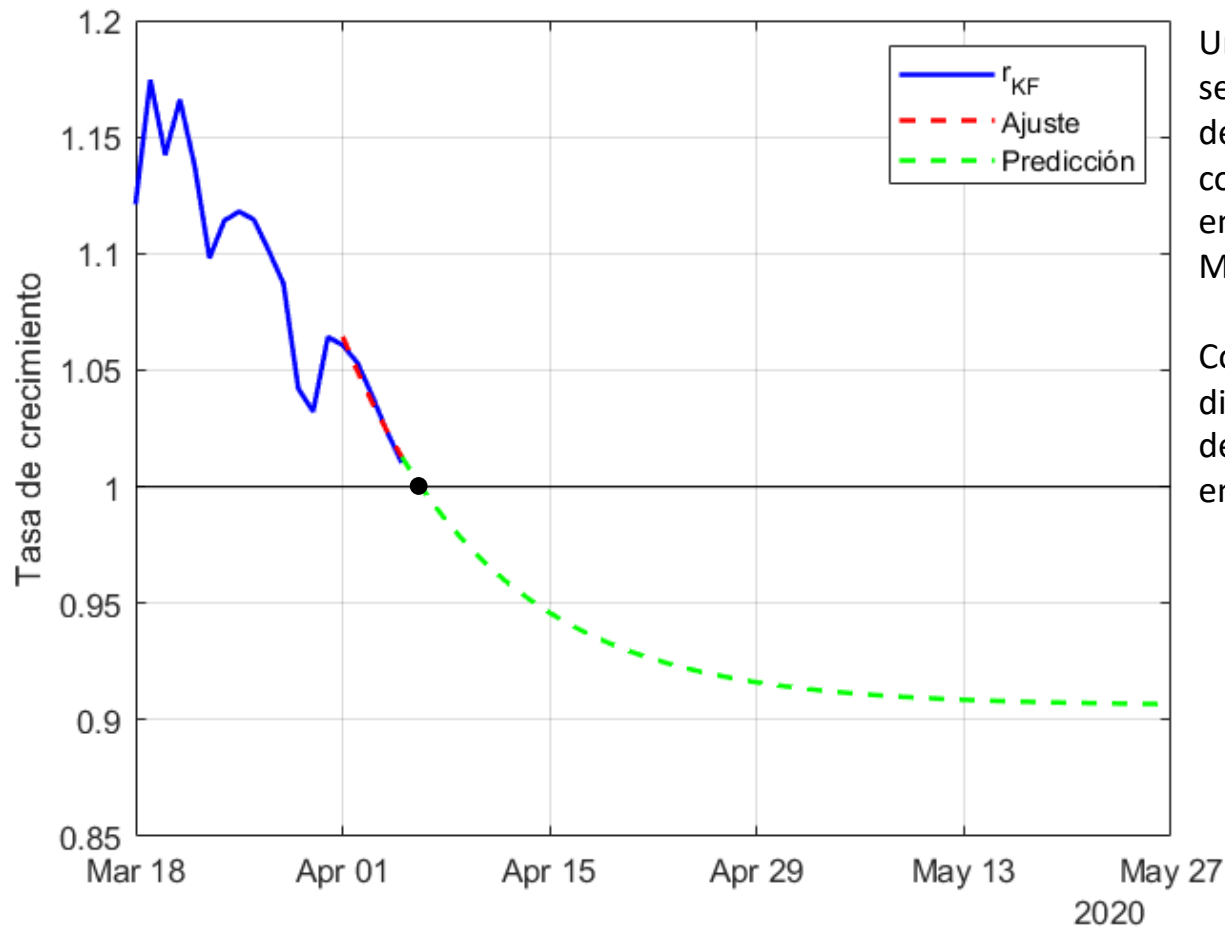
Predicción del número de infectados activos y positivos reportados en Cataluña



Con las mismas consideraciones en lo que concierne al factor k , se muestra en la gráfica adjunta la predicción del número de infectados en la comunidad catalana.

Conjuntamente a estas curvas se ha añadido también la predicción del número de positivos activos reportados.

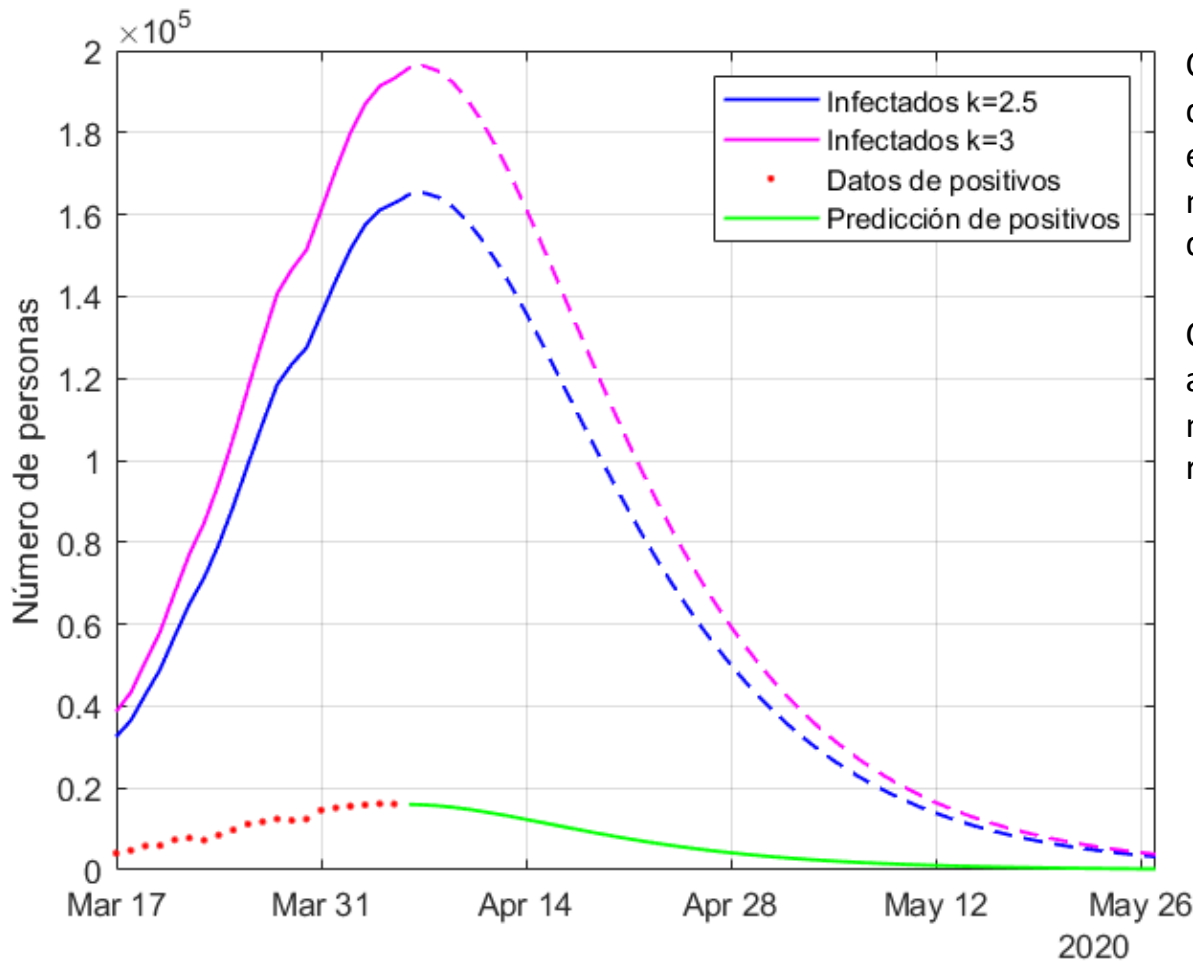
Análisis de la tasa de crecimiento en Madrid



Un procedimiento análogo al de España se ha empleado para la extrapolación de la tasa $r(n)$ en diversas comunidades autónomas, mostrándose en la figura el caso de la comunidad de Madrid.

Con una ventana de los últimos 5 datos disponibles, se prevé que el paso por 1 de la tasa de crecimiento se produzca entre el 8 y el 9 de abril.

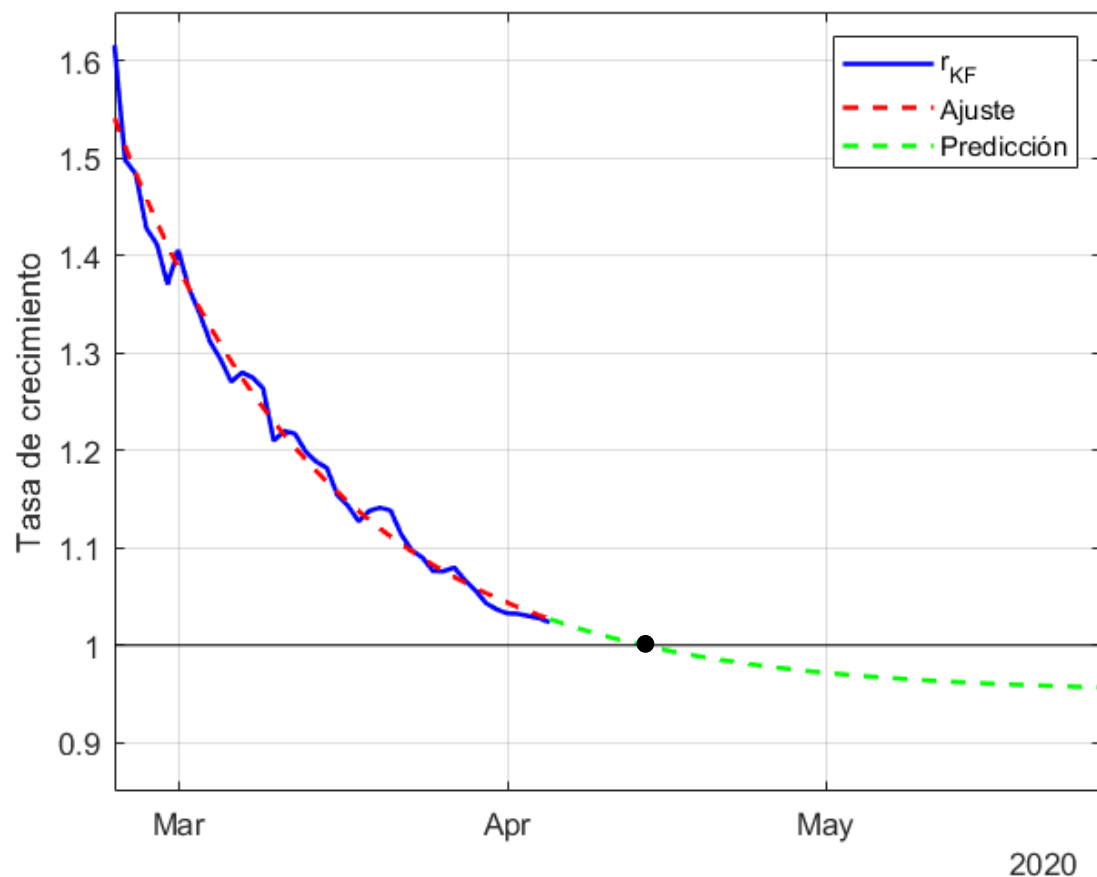
Predicción del número de infectados activos y positivos reportados en Madrid



Con las mismas consideraciones en lo que concierne al factor k , se muestra en la gráfica adjunta la predicción del número de infectados en la comunidad de Madrid.

Conjuntamente a estas curvas se ha añadido también la predicción del número de positivos activos reportados.

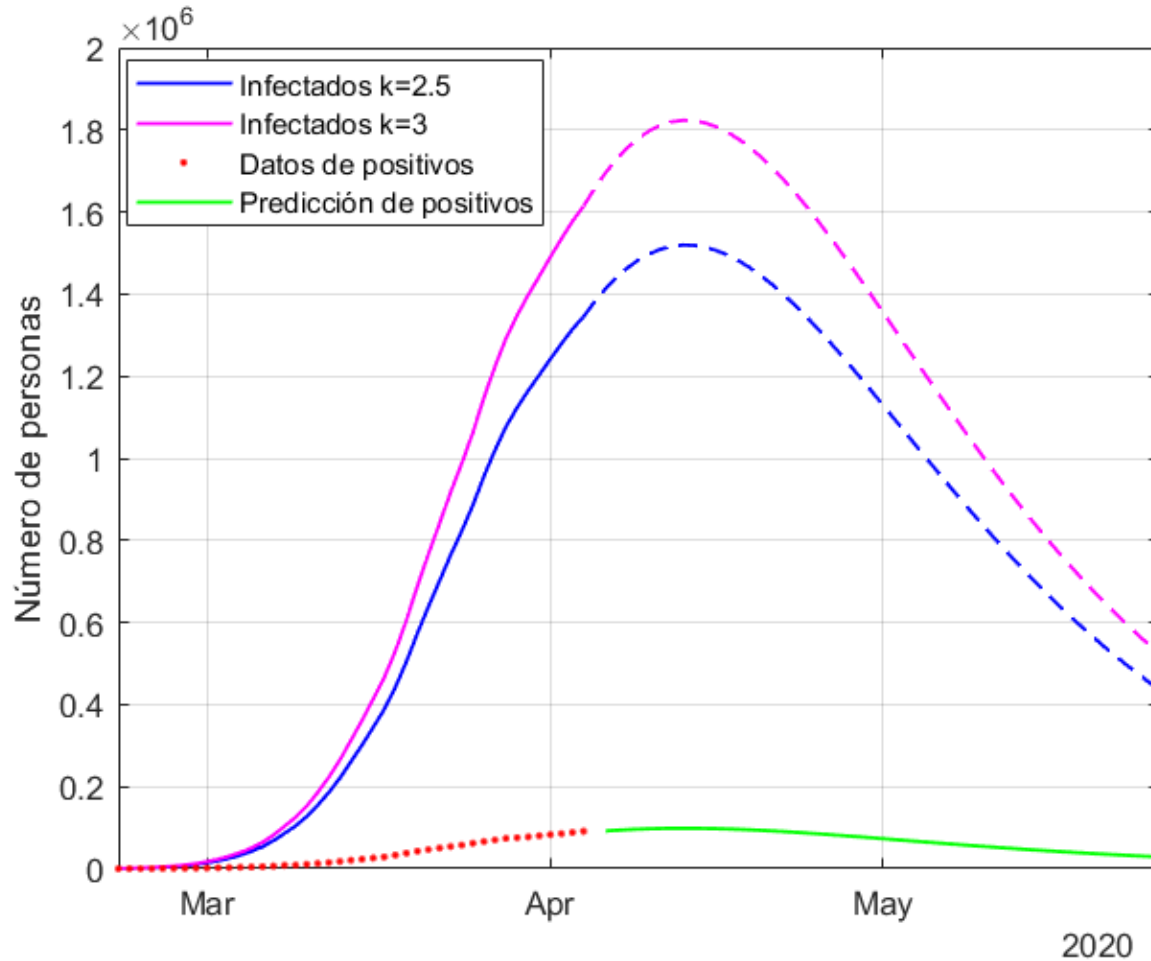
Análisis de la tasa de crecimiento en otros países. Italia



Un procedimiento análogo al de España se ha empleado para la extrapolación de la tasa $r(n)$ en otros países, mostrándose en la figura el caso de Italia.

En este país, tomando todos los datos disponibles hasta la fecha para realizar la extrapolación exponencial, se prevé que el paso por 1 de la tasa de crecimiento se produzca entre el 9 y el 11 de abril.

Predicción del número de infectados activos y positivos reportados en otros países. Italia

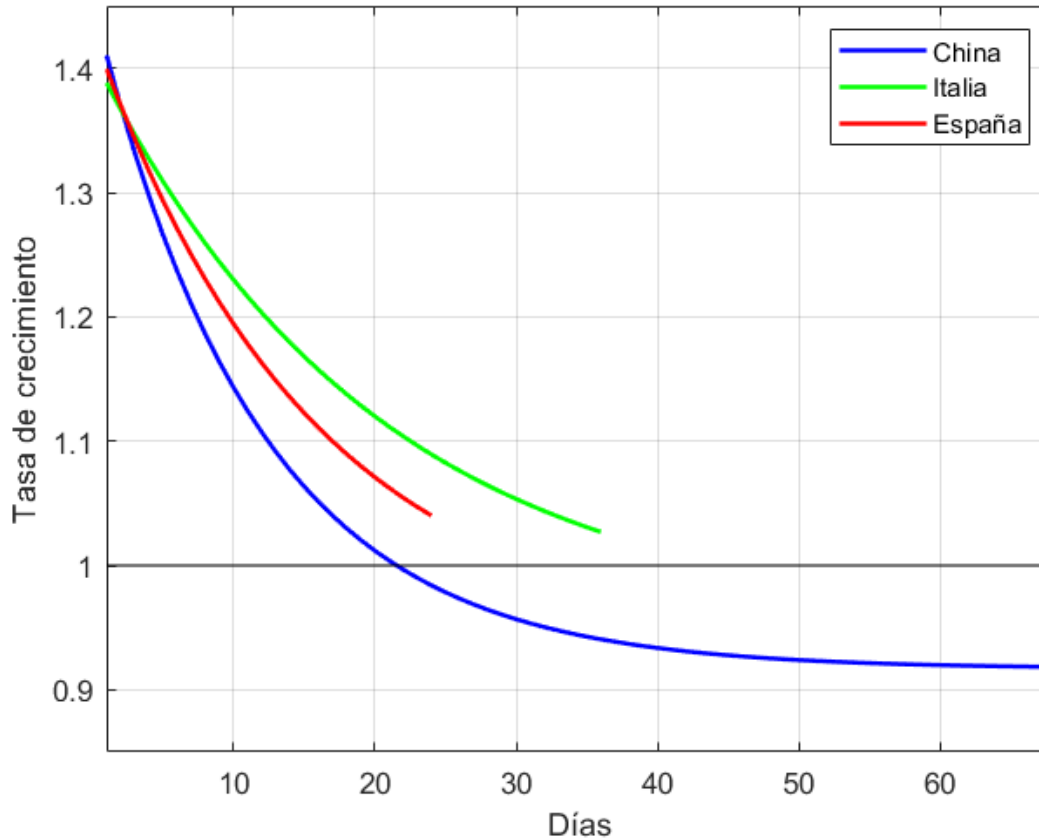


Con las mismas consideraciones tomadas en España y sus comunidades en lo que concierne al factor k , se muestra en la gráfica adjunta la predicción del número de infectados en Italia.

Puede verse como la curva de bajada es más lenta en este país si se compara con la mostrada anteriormente para el territorio español.

Conjuntamente a estas curvas se ha añadido también la predicción del número de positivos activos reportados.

Comparación de la evolución de la tasa de crecimiento en diferentes países



Un último análisis realizado consiste en comparar cómo es la evolución de la tasa de crecimiento en los distintos países considerados en este informe.

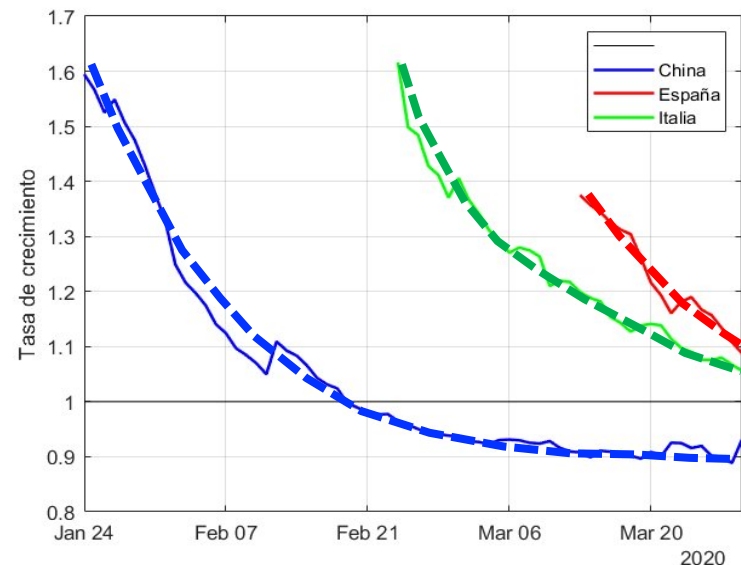
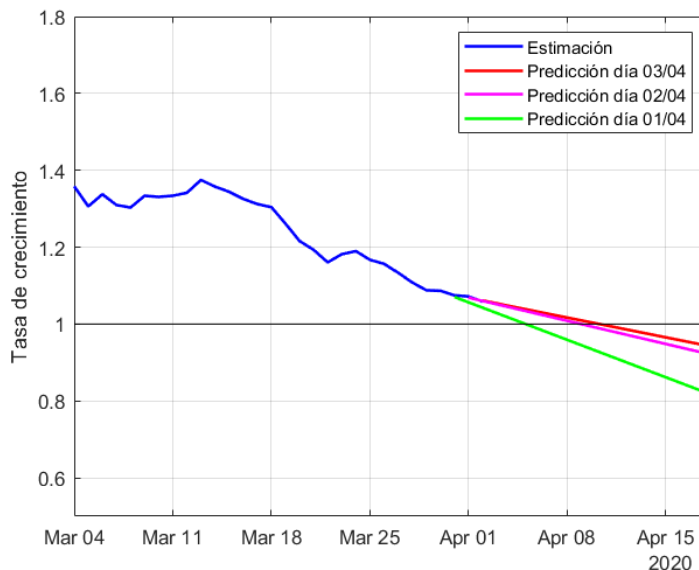
Para ello, se ha cambiado el origen de tiempos de forma que todas las exponenciales estimadas pasen en el mismo instante por un determinado valor característico del periodo anterior al confinamiento (en este caso 1,37, lo cual significa que el número de afectados estaba creciendo el 37% de un día al siguiente).

Puede verse en la gráfica que la constante de tiempo es menor en China, seguida por España e Italia, pudiendo atribuirse este resultado a la dureza del confinamiento en los distintos países.

ANEXO: Caracterización matemática de la evolución de la tasa de crecimiento ante un cambio masivo repentino en la “distancia social”

Como casi todas las aproximaciones a este problema, el análisis comprende dos fases: 1) estimación de la evolución de la epidemia hasta el día actual, en nuestro caso basada en un filtro de Kalman, que fue explicada en el documento inicial y no es objeto de este análisis; 2) predicción de su evolución a futuro, caracterizada a su vez por la evolución de la razón de la progresión geométrica, $r(n)$, que define la tasa de crecimiento ($r(n) > 1$) o decrecimiento una vez pasado el pico ($r(n) < 1$). Este anexo presenta una novedosa justificación teórica, en base a la cual se hará la predicción de la evolución de $r(n)$ en lo sucesivo.

Hasta ahora, la predicción a futuro de $r(n)$ se ha realizado mediante un ajuste lineal sobre una ventana de datos de los últimos días (extrapolación lineal), tal como muestra la figura de la izquierda para el caso de España. Sin embargo, tal como sugieren las líneas a trazos de la gráfica de la derecha, que muestra la evolución de $r(n)$ en tres países, cuando se considera una ventana suficientemente grande, la evolución de $r(n)$ parece ser más bien de tipo exponencial, similar a la respuesta de un sistema de primer orden.



ANEXO: Caracterización matemática de la evolución de la tasa de crecimiento ante un cambio masivo repentino en la “distancia social”

A continuación, vamos a intentar obtener una expresión analítica de la evolución de $r(n)$, tras una perturbación que modifique bruscamente el distanciamiento social (y por tanto el número reproductivo básico del virus, R_0), tal como ocurre con el confinamiento impuesto en numerosos países.

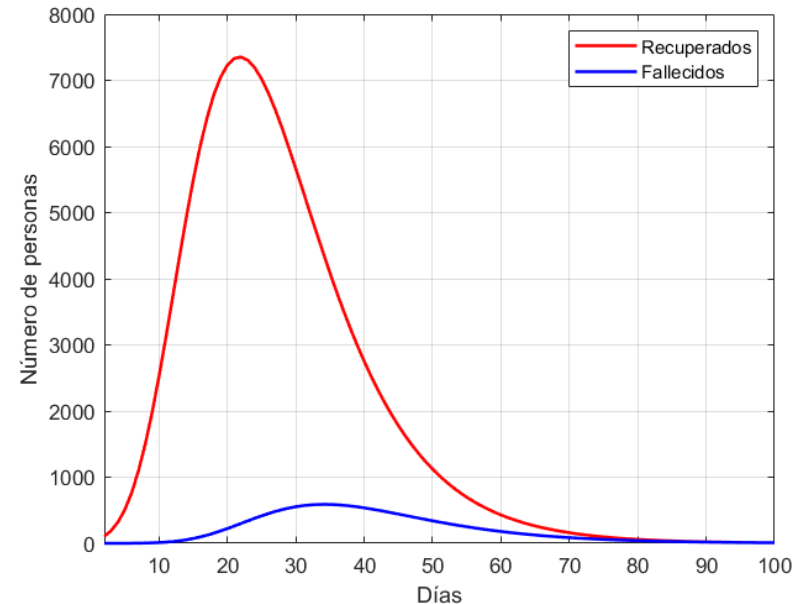
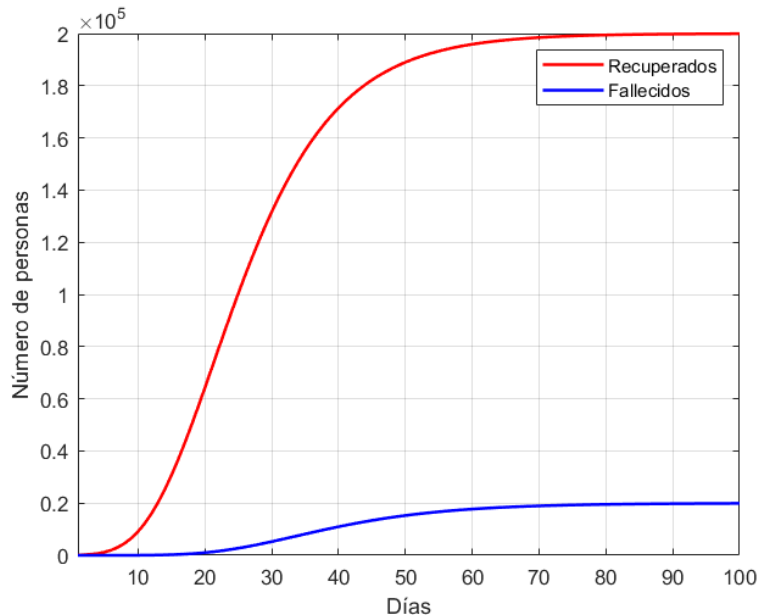
Conviene en este punto aclarar una vez más que, aunque están relacionadas, no debe confundirse la constante $r(n)$ de la progresión geométrica de infectados con la constante R_0 , conocida como número reproductivo básico (número medio de personas que cada infectado contagia cuando la sociedad no toma ninguna medida), que depende de cada virus y contexto socioeconómico (para el Covid-19 se estima que R_0 está comprendida entre 2 y 3). A partir del modelo SIR, se deduce la siguiente relación aproximada entre ambas constantes: $r(n) \approx 1 + (R_0 - 1)/T$, donde T es el periodo medio infeccioso. Por ejemplo, para $R_0 = 2,9$ y un periodo infeccioso de 10 días, resulta $r(n) \approx 1,19$.

Antes de proseguir, es necesario también insistir en las notables diferencias conceptuales y de modelo que resultan cuando se trabaja con cifras acumulativas o cifras de nuevos casos diarios. Las primeras nunca decrecen, por definición, y responden a una curva sigmoidea. Las segundas se obtienen por diferencia entre cifras acumuladas consecutivas (proceso matemático de derivación), y tienen una forma acampanada. De esta forma, podemos registrar fallecidos diarios o fallecidos acumulados (suma o integración de diarios), y lo mismo con los recuperados.

En cuanto a los positivos confirmados (en algunos medios llamados ambiguamente “contagiados”) podemos distinguir entre la cifra total y la de “activos”, esta última obtenida restando del total de positivos tanto los fallecidos como los recuperados acumulados, que están lógicamente retrasados en el tiempo. Lo propio puede decirse de los infectados reales (necesariamente estimados), que se reducen a los infectados activos (incluyendo asintomáticos) cuando se deducen los recuperados y fallecidos.

ANEXO: Caracterización matemática de la evolución de la tasa de crecimiento ante un cambio masivo repentino en la “distancia social”

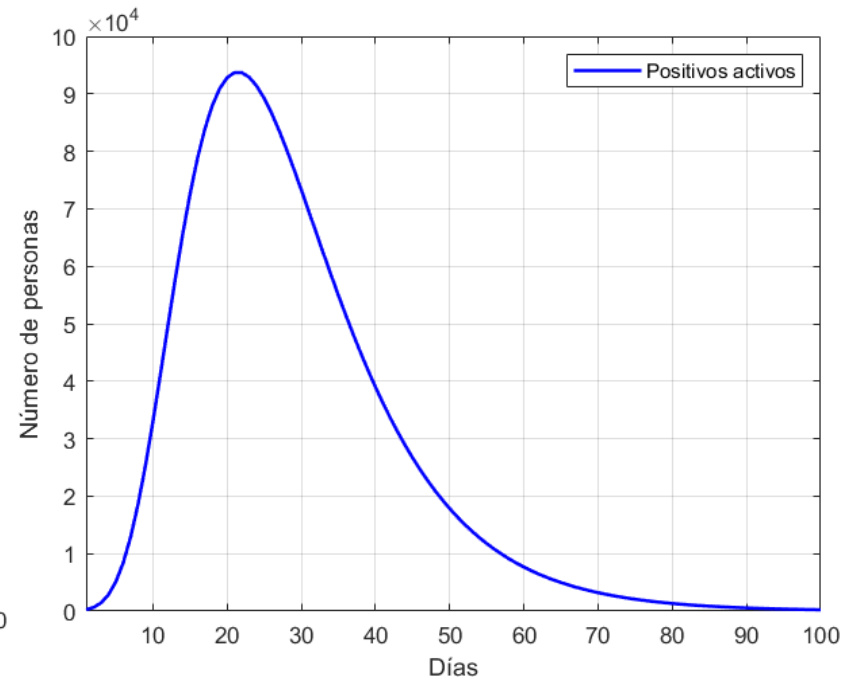
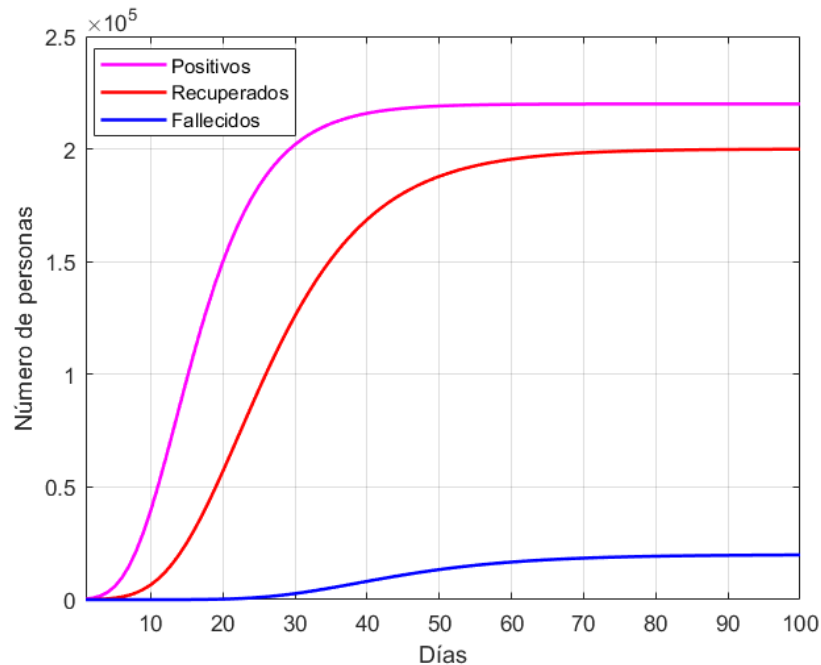
La figura de la izquierda muestra sendos gráficos idealizados con los datos de fallecidos y recuperados acumulados, mientras que la figura de la derecha muestra el registro diario de ambas magnitudes, obtenidas por diferencia entre dos días consecutivos (derivación). Los gráficos reales serían más irregulares, debido a errores en la recopilación de datos, pero las curvas de acumulados siempre serían crecientes, hasta alcanzar un máximo asintótico, mientras que las curvas de datos diarios tenderían asintóticamente a cero, suponiendo que se llega a un estado estacionario y se supera la enfermedad.



Resulta evidente que el valor de $r(n)$ tiende a la unidad para cualquier curva acumulativa, puesto que, conforme se alcanza la asíntota, $x(n) \approx x(n - 1)$, mientras que $r(n)$ está en principio indeterminado en el caso de la curva de datos diarios ($0/0$), puesto que, en el límite, $x(n) \approx 0$ y $x(n - 1) \approx 0$.

ANEXO: Caracterización matemática de la evolución de la tasa de crecimiento ante un cambio masivo repentino en la “distancia social”

La figura de la izquierda muestra la evolución típica del número de positivos, fallecidos y recuperados acumulados, mientras que la de la derecha muestra el saldo neto de positivos activos. El que esta última magnitud sea la diferencia entre curvas acumulativas, y el que los positivos totales iguale asintóticamente la suma de fallecidos y recuperados, explica que la curva de positivos activos sea también en forma de campana, con tendencia asintótica a cero.



Idéntico razonamiento podría aplicarse a la curva de infectados activos, con la diferencia de que el balance incluye en este caso no sólo los positivos confirmados, sino toda la población contagiada (asintomáticos o que no han sido sometidos al test).

ANEXO: Caracterización matemática de la evolución de la tasa de crecimiento ante un cambio masivo repentino en la “distancia social”

Las curvas sigmoideas que representan datos acumulativos pueden aproximarse de forma analítica mediante diferentes funciones, siendo una de las más usadas para modelar epidemias víricas la llamada función de Gompertz. Dada una variable acumulativa $x_a(t)$, dicha función sigmoidea viene dada por las siguientes expresiones exponenciales acopladas, donde K y a son constantes a determinar en cada caso y x_0 es el valor de la variable en el instante inicial, t_0 :

$$x_a(t) = Ke^{y(t)} \quad y(t) = y_0 e^{-a(t-t_0)} \quad y_0 = \ln\left[\frac{x_0}{K}\right]$$

Alternativamente, la curva de Gompertz puede reescribirse como un sistema de dos ecuaciones diferenciales de primer orden, una de ellas no lineal:

$$\dot{x}(t) = -ax(t)y(t) \quad \dot{y}(t) = -ay(t)$$

Puede demostrarse que, si la función $x_a(t)$ se representa como una serie geométrica de razón variable, es decir, $x_a(t) = r_a(t) x_a(t - \Delta T)$, donde ΔT es el intervalo de muestreo (en este caso un día), la razón geométrica viene dada por la expresión:

$$r_a(t) = 1 - a\Delta T y_0 e^{-a(t-t_0)}$$

Como puede comprobarse, el valor $r_a(t)$ de una magnitud acumulativa $x_a(t)$ tiende a la unidad asintóticamente, como se espera que ocurra en una curva sigmoidea.

ANEXO: Caracterización matemática de la evolución de la tasa de crecimiento ante un cambio masivo repentino en la “distancia social”

Una vez obtenida analíticamente la razón geométrica $r_a(t)$ de una serie de datos acumulativos $x_a(t)$, puede deducirse también algebraicamente la razón geométrica $r_d(t)$ de la correspondiente serie de datos diarios, $x_d(t)$, obtenida a partir de la serie acumulativa original. Teniendo en cuenta que la serie diaria se obtiene por diferenciación (derivación) de la serie acumulativa ($x_d(t)=\dot{x}_a(t)$), puede demostrarse que:

$$r_d(t) = e^{-a \Delta T} - a \Delta T y_0 e^{-a (t-t_0)}$$

Obsérvese que, tanto $r_a(t)$ como $r_d(t)$ evolucionan exponencialmente con la misma constante de tiempo ($T_r = 1/a$), pero tienden asintóticamente a valores diferentes. Mientras que $r_a(t)$ tiende a 1, de acuerdo a lo esperado teóricamente, $r_d(t)$ tiende a un valor que, a diferencia de lo que ocurre en un sistema de primer orden ordinario, depende de la constante de tiempo: $r_d(\infty) = e^{-a \Delta T}$. Por ejemplo, si la constante de tiempo fuese 10 días, $r_d(t)$ tendería a 0,9. Las siguientes figuras muestran la evolución de ambas razones geométricas en un caso hipotético, suponiendo que la curva acumulativa responde a la forma ideal de Gompertz.

