

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, que en principio será el último, muestra los resultados obtenidos con los datos disponibles a 12 de abril de 2020. Se remite al lector interesado a la serie de documentos publicados desde el 25 de marzo, que se pueden descargar de: <https://idus.us.es/handle/11441/94508>

Resumen

En el informe de hoy, último de la serie, además de actualizar los resultados de España, las comunidades autónomas, y resto de países, se explica en qué medida podría utilizarse el método propuesto durante la fase de “desescalamiento” de la enfermedad a la que estamos abocados desde hoy. Así mismo, se incluyen una serie de reflexiones, a modo de epílogo, resaltando las aportaciones realizadas.

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.

Agradecimientos: Agradecemos la ayuda, consejos y permanente disponibilidad de Consuelo Arahal y Mercedes Aguilar, para actualizar en idUS esta serie de documentos.

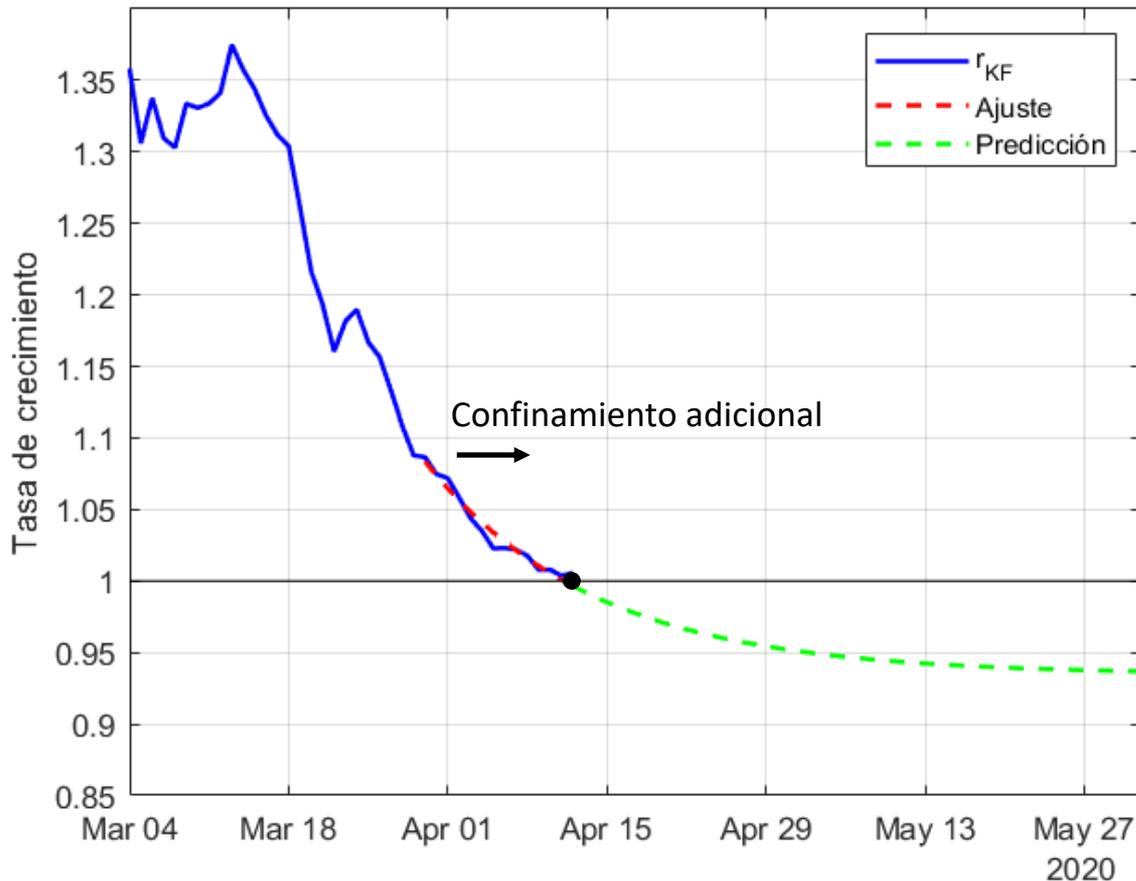


Esta obra está bajo una [licencia de Creative Commons Reconocimiento 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).

Contenido

	Pag.
• Resultados actualizados para España _____	3
• Resultados actualizados para CCAA seleccionadas _____	6
• Resultados actualizados para otros países _____	12
• Resumen de resultados para España: serie histórica _____	22
• Utilidad del modelo desarrollado en la fase de remisión ____	24
• Epílogo _____	27

Análisis de la tasa de crecimiento en España

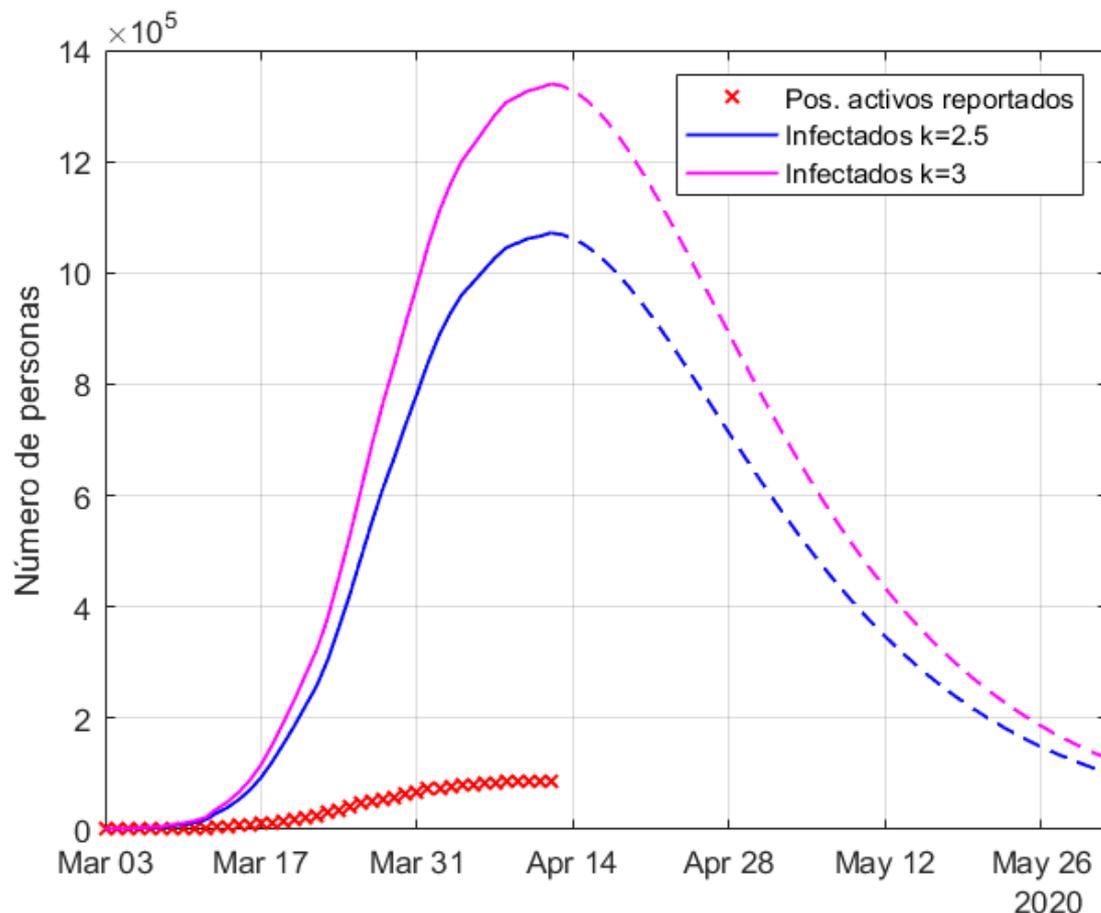


El método propuesto se basa en la obtención de un ajuste exponencial que permita realizar una extrapolación de $r(n)$, habiéndose validado este procedimiento con el caso de China, como se mostró en el informe del día 6 de abril. Se han tomado los datos disponibles desde el decreto de confinamiento adicional el día 31 de marzo para el cálculo del mencionado ajuste.

El resultado, mostrado en la figura, conduce a un paso por 1 (pico) muy próximo en el tiempo a la fecha en que se redacta este informe. Además, se observa un valor asintótico de la tasa de crecimiento próximo a 0,93 (ligeramente superior al de otros países como China).

Este ajuste será utilizado para la obtención de los datos que se van a mostrar a continuación.

Predicción del número de infectados activos en España



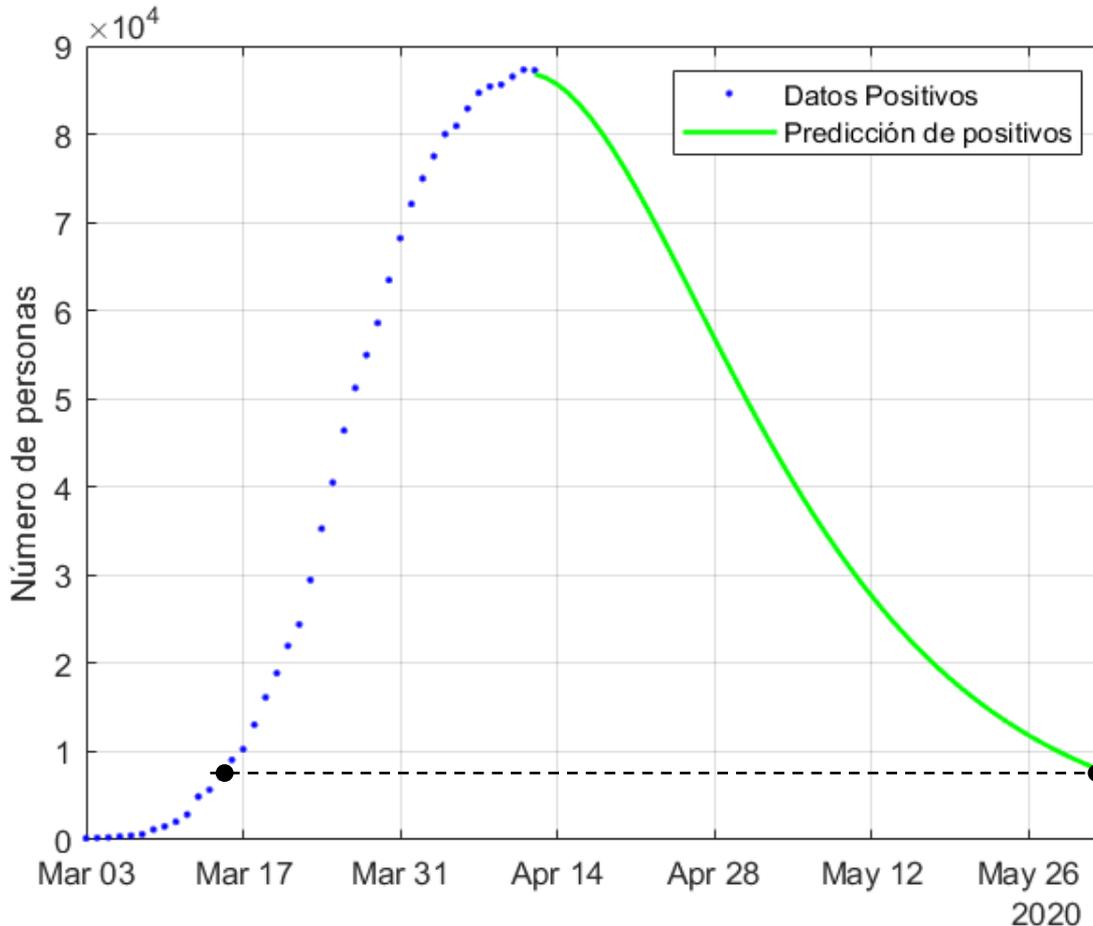
Valores estimados en línea continua y predicciones en discontinua

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 1,35 millones de personas, para los números iniciales considerados.

Predicción del número de positivos activos reportados en España

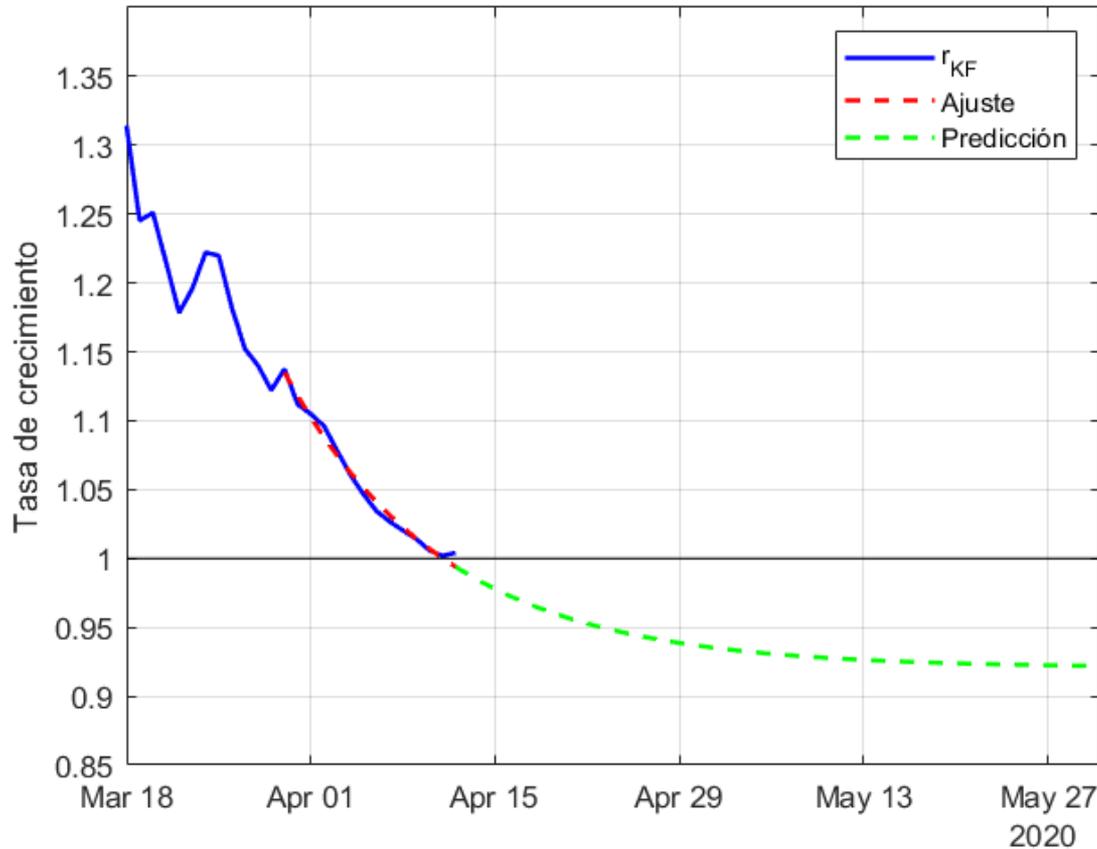


Finalmente, se muestra para el territorio español la predicción del número de positivos activos reportados, pudiendo verse que, en el pico, el número de positivos activos está entre 85 y 90 mil personas.

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 30 de mayo (como sugiere la línea discontinua), suponiendo que no se modifica el confinamiento actual.

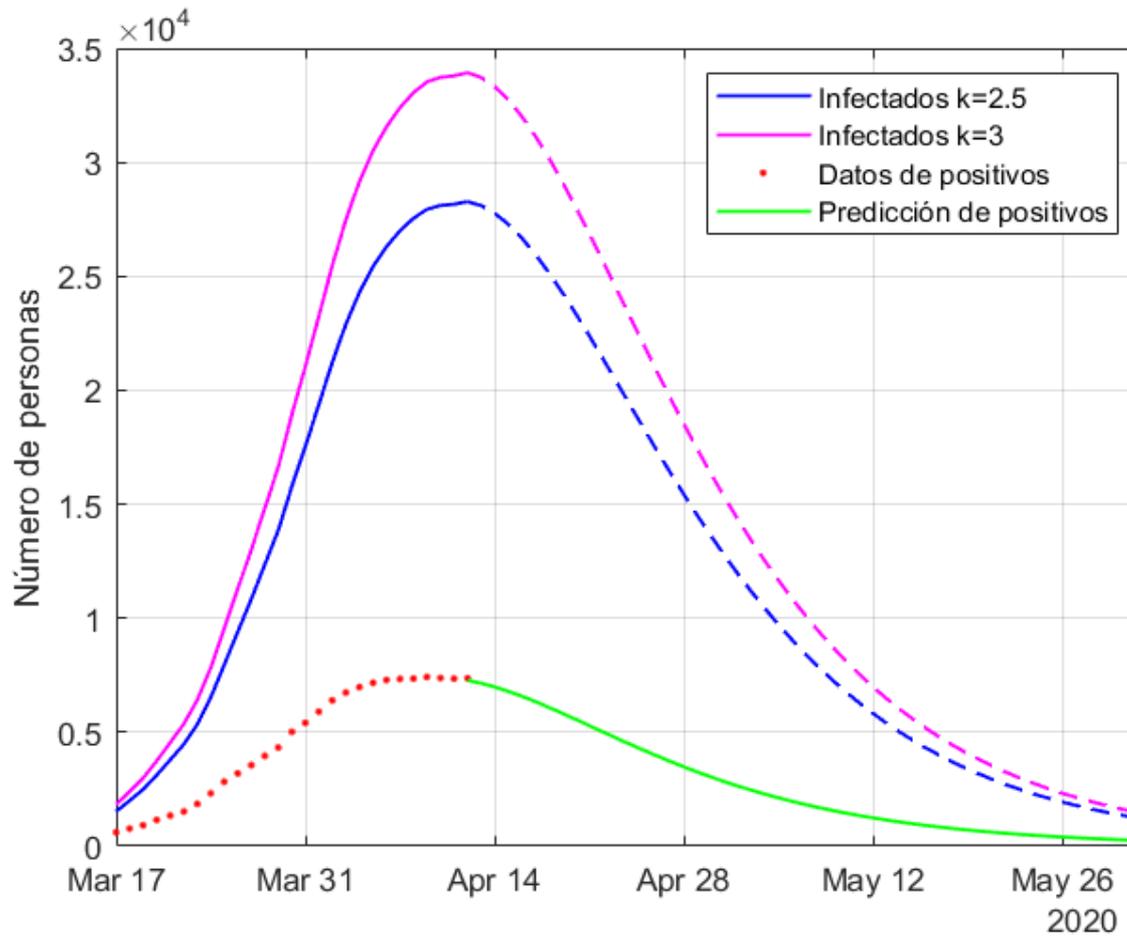
Análisis de la tasa de crecimiento en Andalucía



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 correspondiente a los datos de las dos últimas semanas, puede comprobarse como en esta comunidad la tasa de crecimiento está muy próxima a la unidad, correspondiéndose con el pico de los infectados activos.

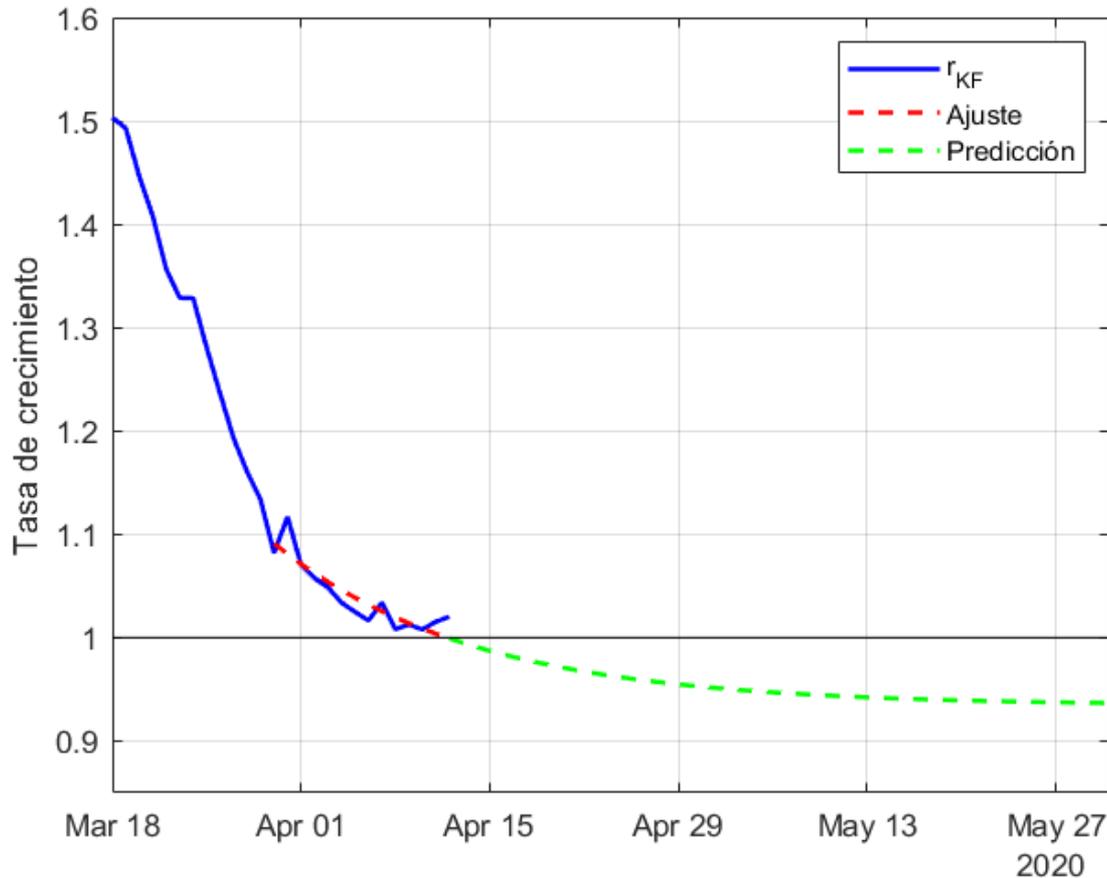
Predicción del número de infectados activos y positivos reportados en Andalucía



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.

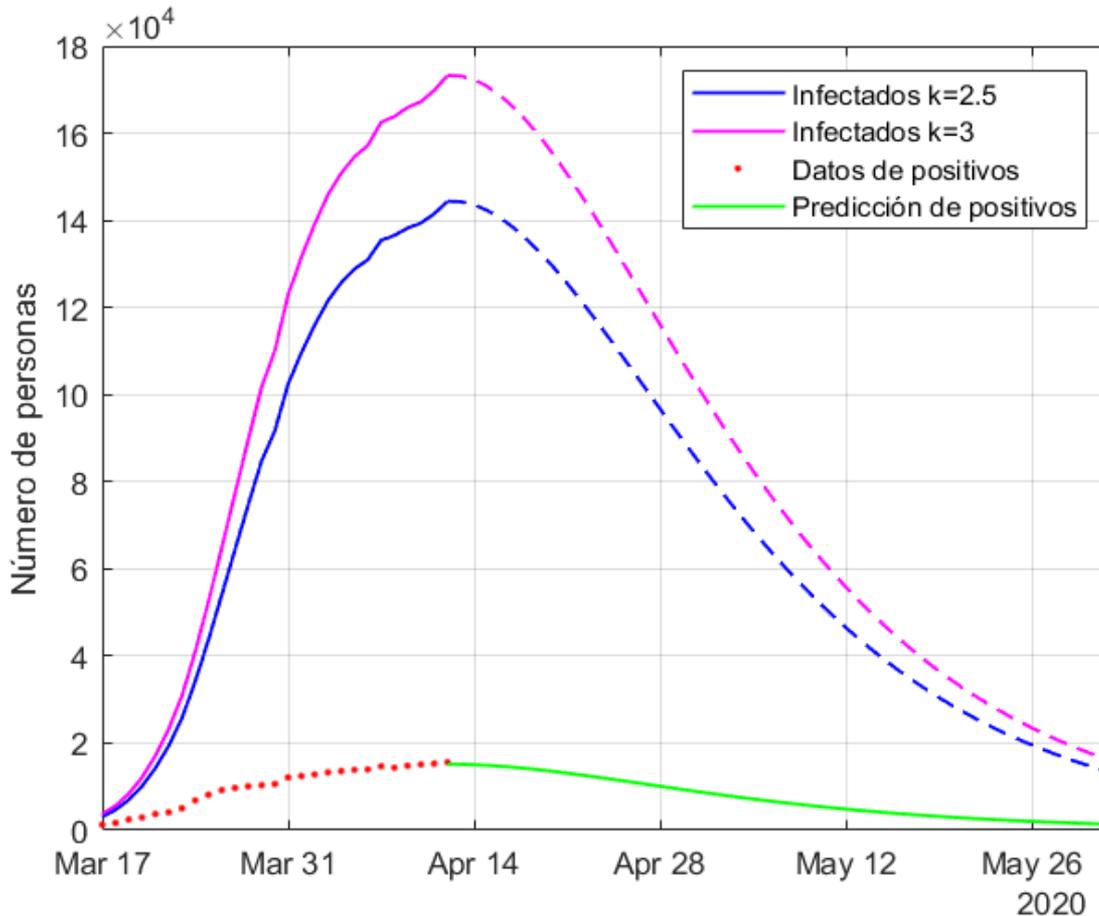
Análisis de la tasa de crecimiento en Cataluña



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 correspondiente a los datos de las dos últimas semanas, puede comprobarse como en esta comunidad la tasa de crecimiento está muy próxima a la unidad, correspondiéndose con el pico de los infectados activos, a pesar de que los últimos reportes de positivos no hayan sido muy favorables.

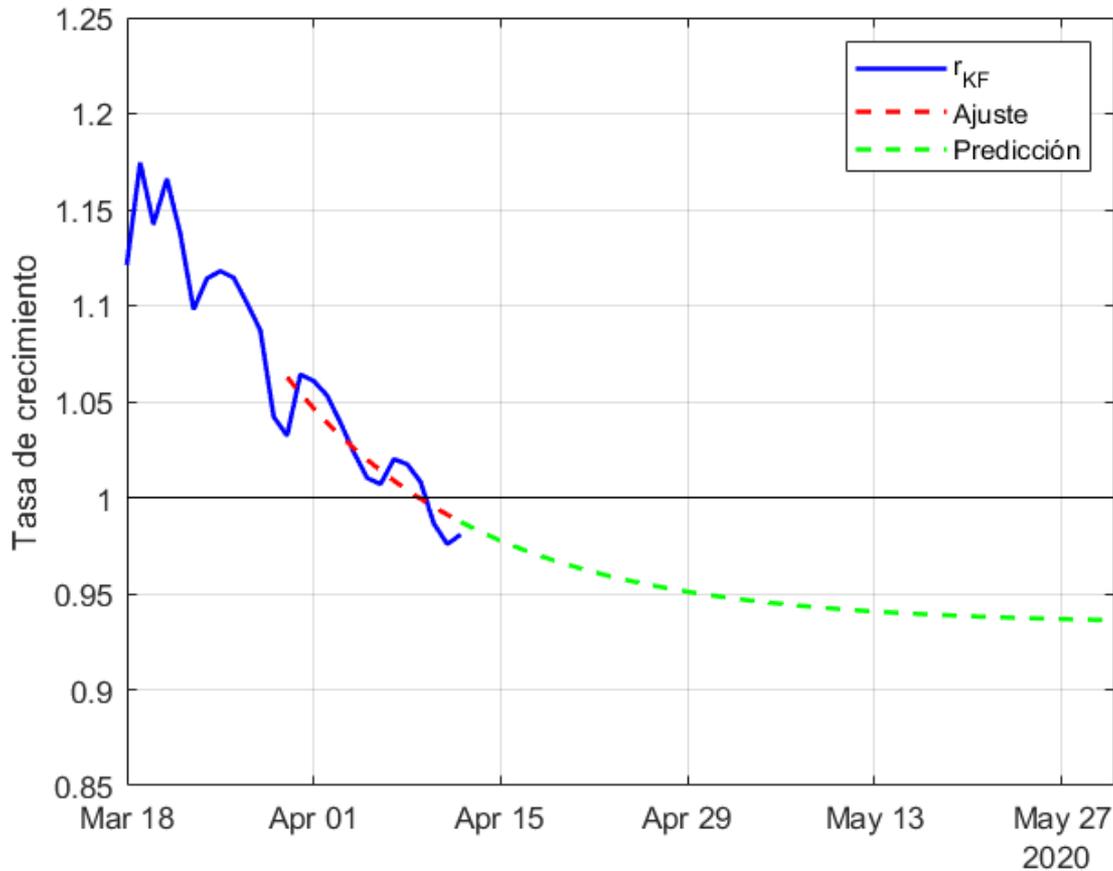
Predicción del número de infectados activos y positivos reportados en Cataluña



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.

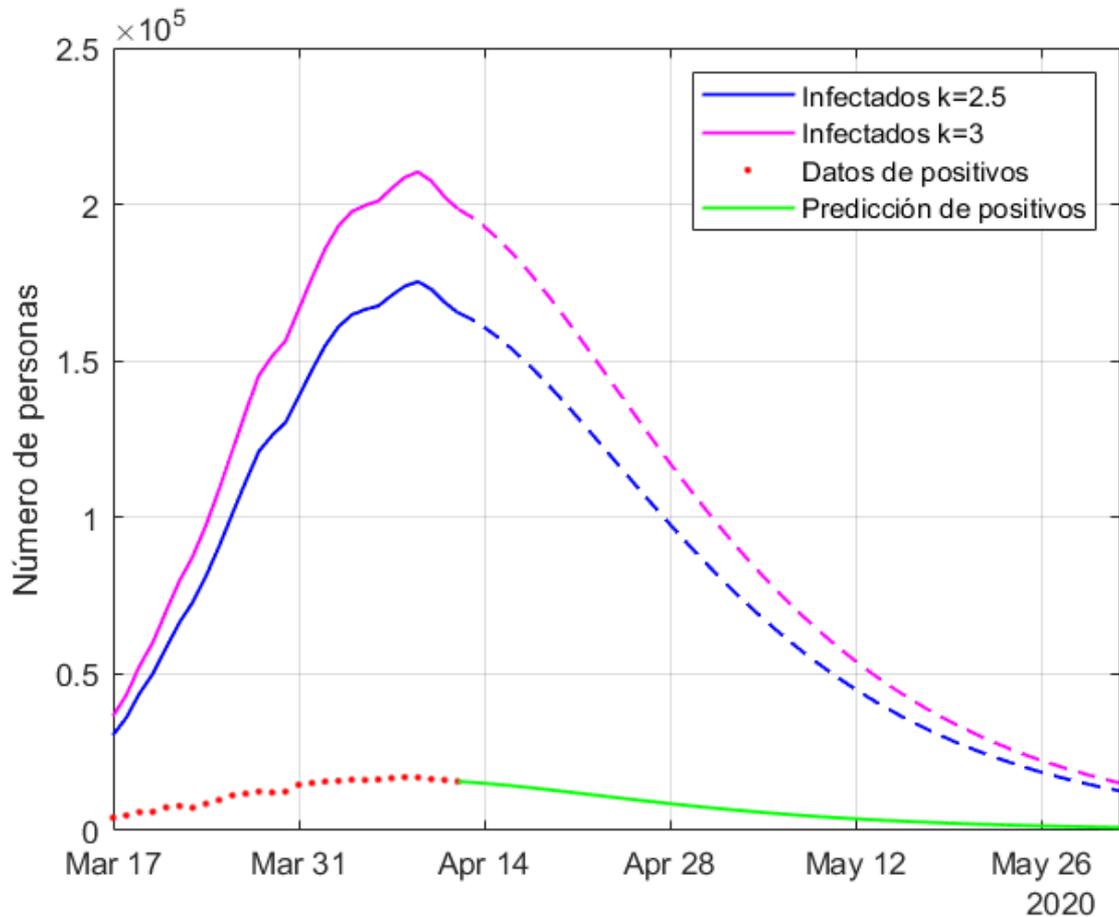
Análisis de la tasa de crecimiento en Madrid



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 correspondiente a los datos de las dos últimas semanas, puede verse como en esta comunidad, la más castigada por la pandemia, ya se sobrepasó hace unos días la barrera unitaria en la tasa de crecimiento, lo cual son buenas noticias.

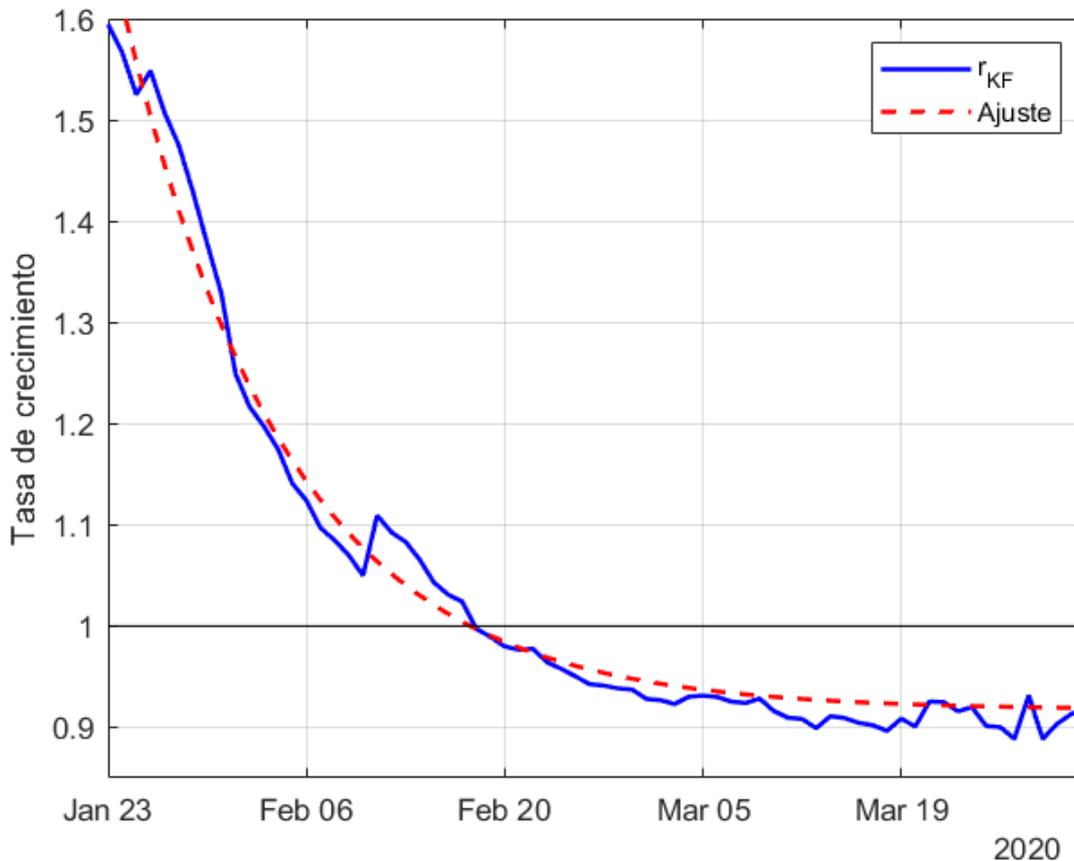
Predicción del número de infectados activos y positivos reportados en Madrid



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.

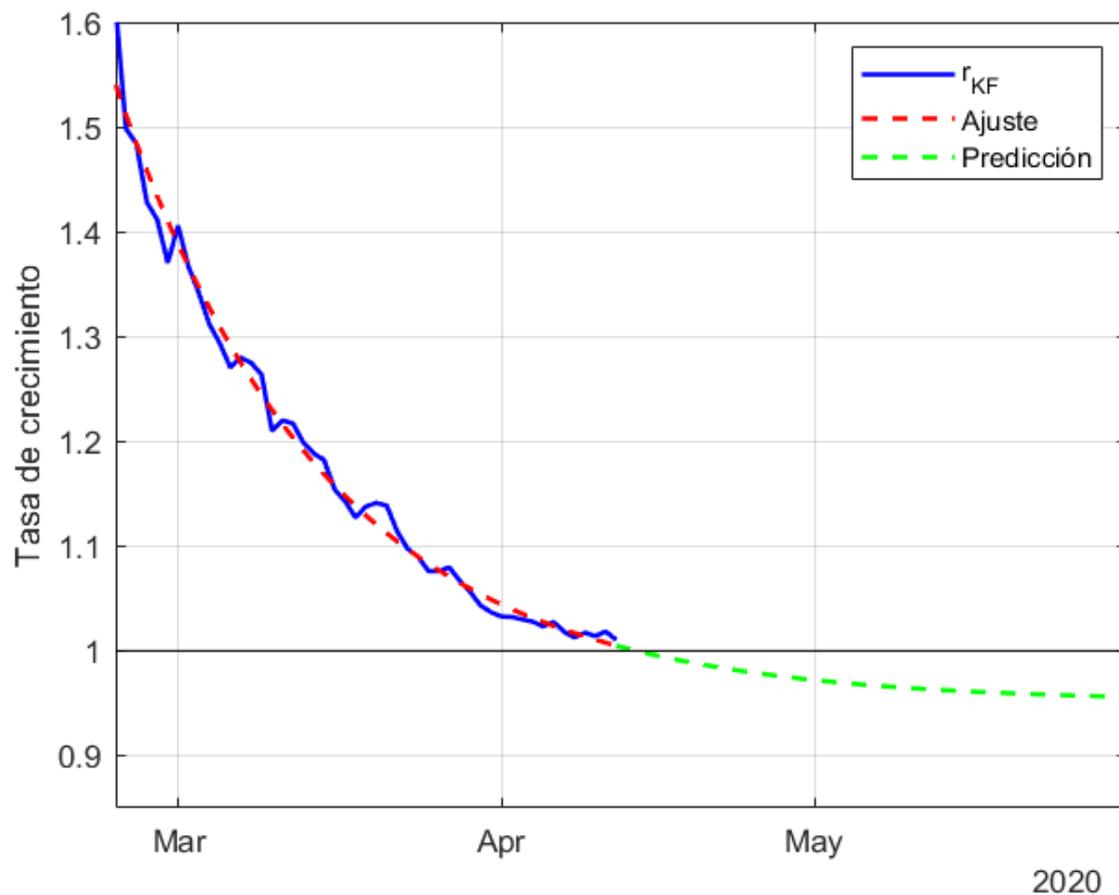
Análisis de la tasa de crecimiento en otros países. China



Dado que en China se inició la pandemia con anterioridad al resto de países, se ha tomado como base para evaluar la eficacia de la estimación y el posterior ajuste exponencial de $r(n)$, que permitirá la extrapolación posterior.

En este país puede comprobarse como la curva exponencial se ajusta de manera correcta a la evolución estimada por el filtro de Kalman de la tasa de crecimiento.

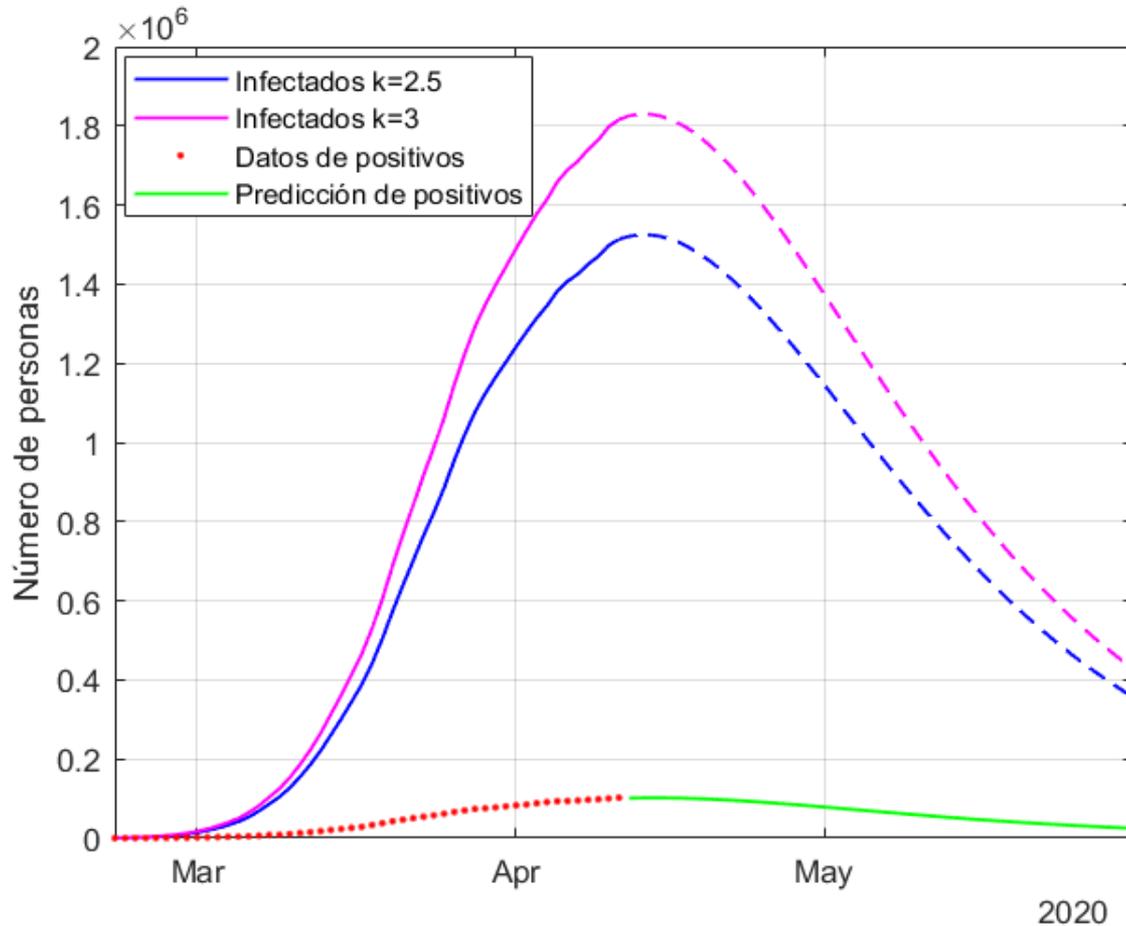
Análisis de la tasa de crecimiento en otros países. Italia



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 con la publicación de los próximos datos de positivos.

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

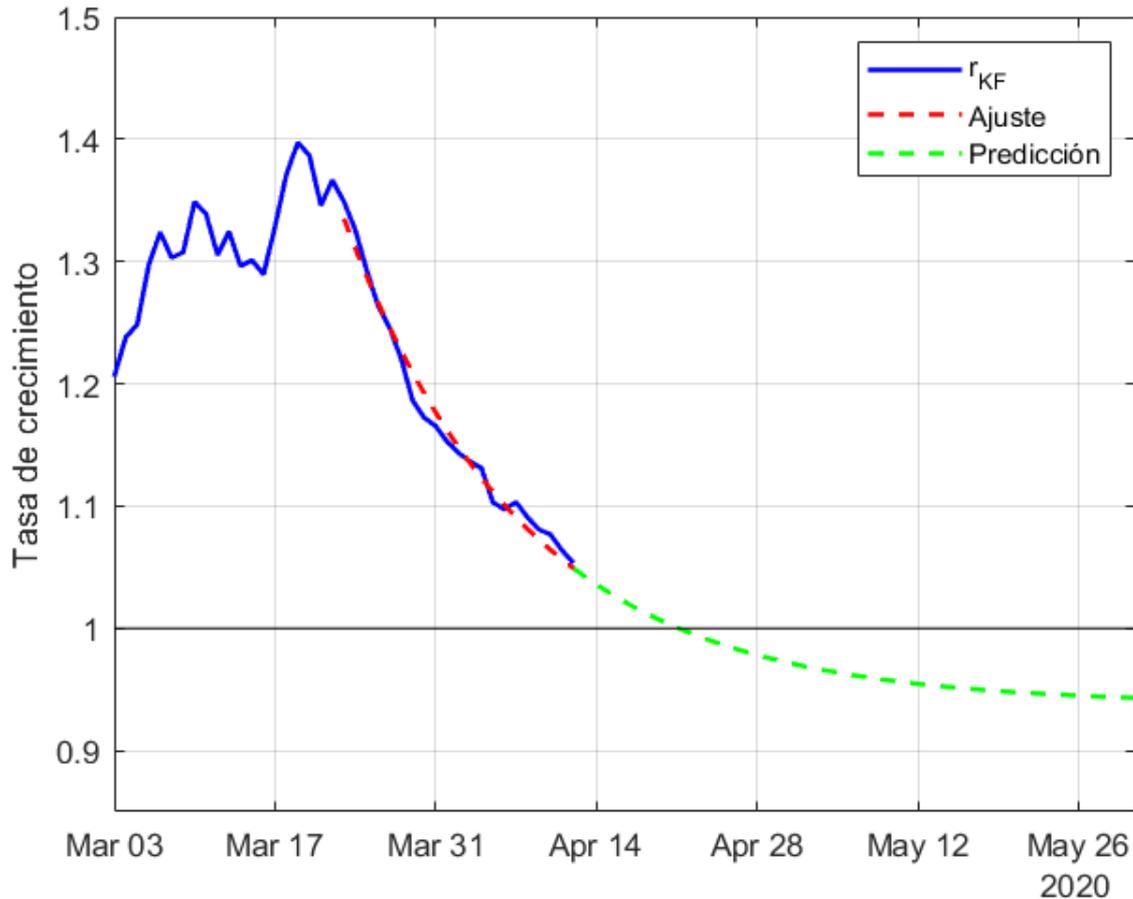


Con las mismas consideraciones tomadas en España y sus comunidades en lo que concierne al factor k , se muestra en la gráfica adjunta la predicción del número de infectados en Italia.

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.

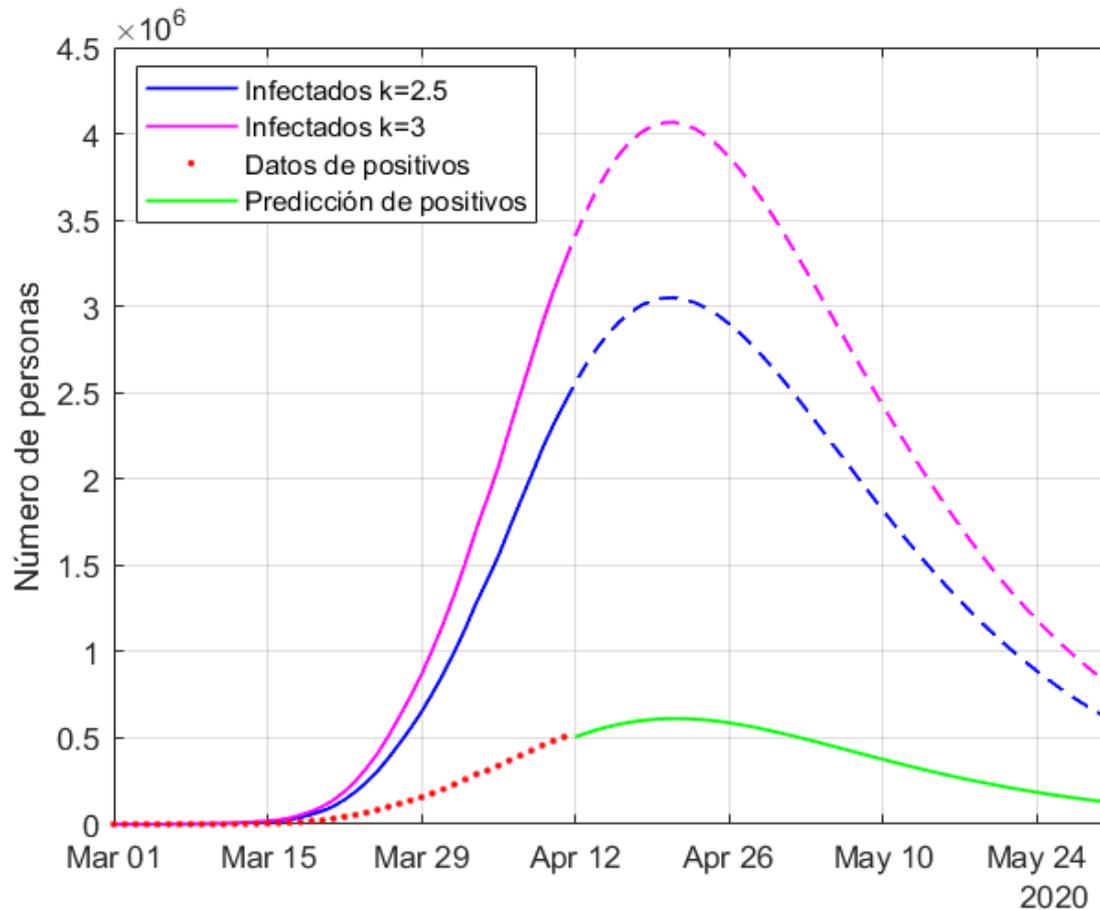
Análisis de la tasa de crecimiento en otros países. Estados Unidos



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 Estados Unidos.

En este país, tomando una ventana de los últimos 20 datos disponibles, se prevé que el paso por 1 de la tasa de crecimiento se produzca entre el 20 y el 22 de abril.

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

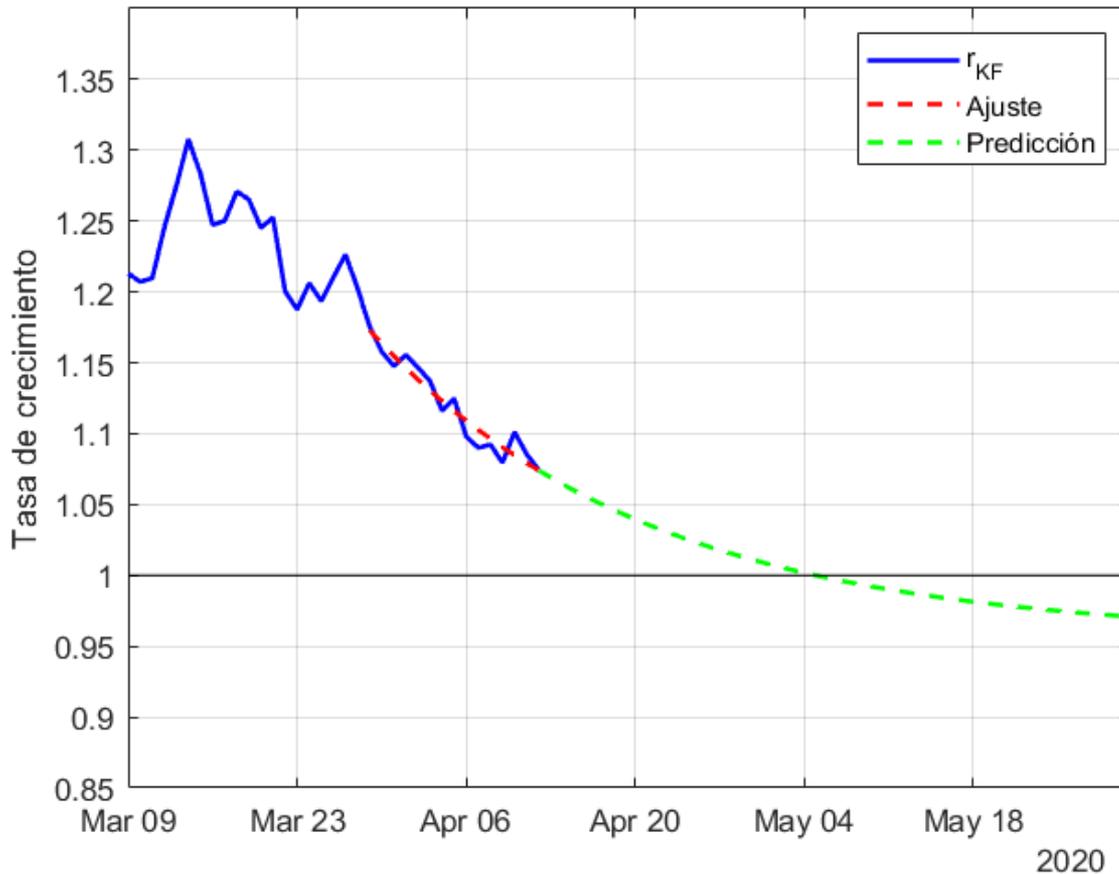


Con las mismas consideraciones tomadas en España y sus comunidades en lo que concierne al factor k , se muestra en la gráfica adjunta la predicción del número de infectados en Estados Unidos.

En este caso se estima que el número máximo de infectados activos estará entre 3 y 4 millones de personas.

Conjuntamente a estas curvas se ha añadido también la predicción del número de positivos activos reportados, que tendrá un pico en torno a 600 mil.

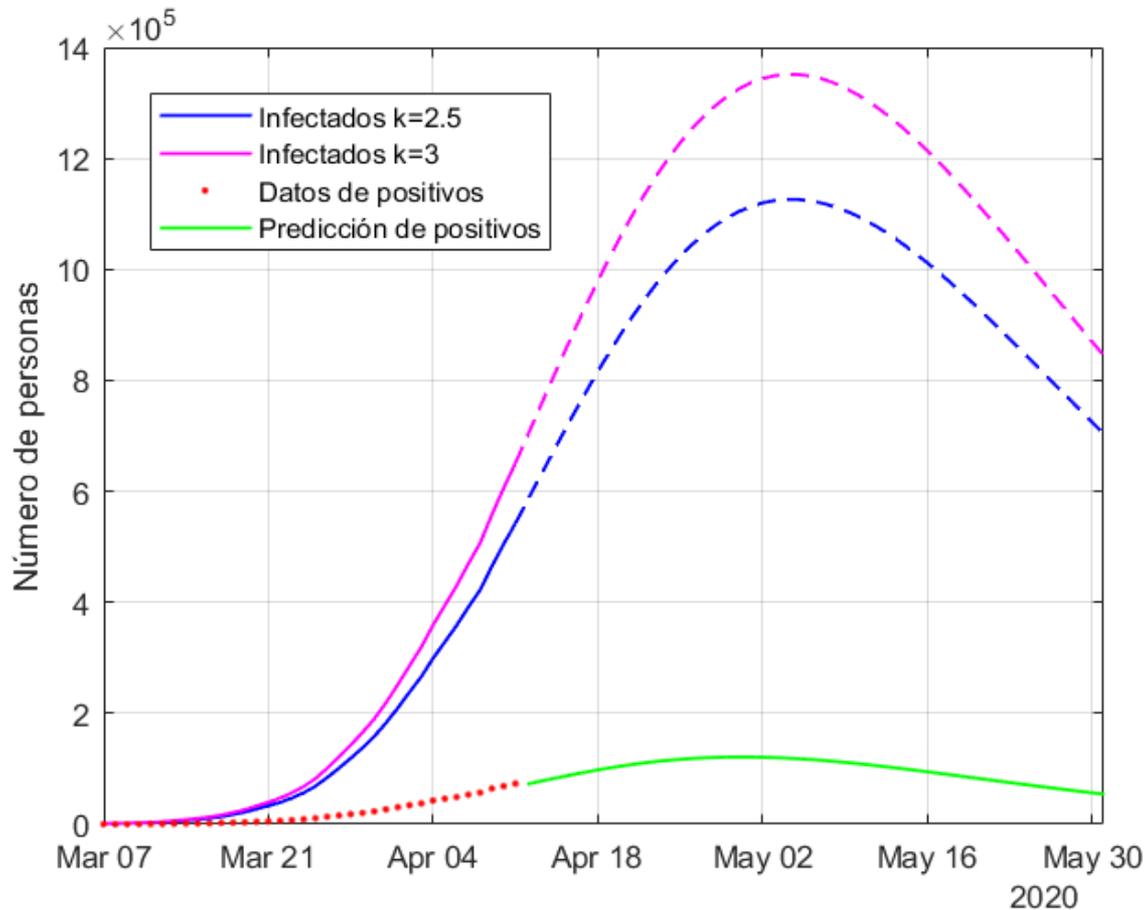
Análisis de la tasa de crecimiento en otros países. Reino Unido



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 Reino Unido.

En este país, tomando una ventana de los últimos 15 datos disponibles, se prevé que el paso por 1 de la tasa de crecimiento se produzca entre el 2 y el 4 de mayo.

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

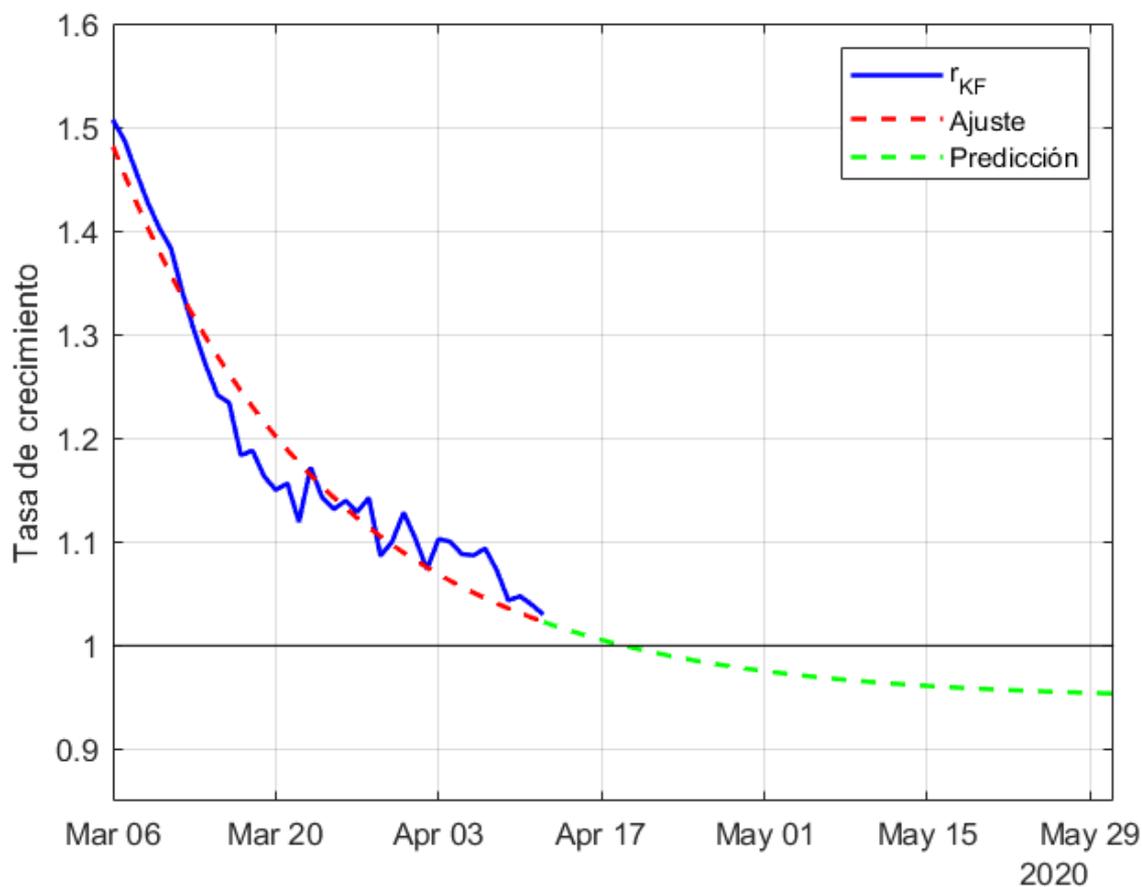


Con las mismas consideraciones tomadas en España y sus comunidades en lo que concierne al factor k , se muestra en la gráfica adjunta la predicción del número de infectados en Reino Unido, tomando una ventana de los últimos 15 datos.

En este caso se comprueba que el número máximo de infectados activos estará entre 1.1 y 1,4 millones de personas.

Conjuntamente a estas curvas se ha añadido también la predicción del número de positivos activos reportados, que tendrá un pico en torno a los 125 mil.

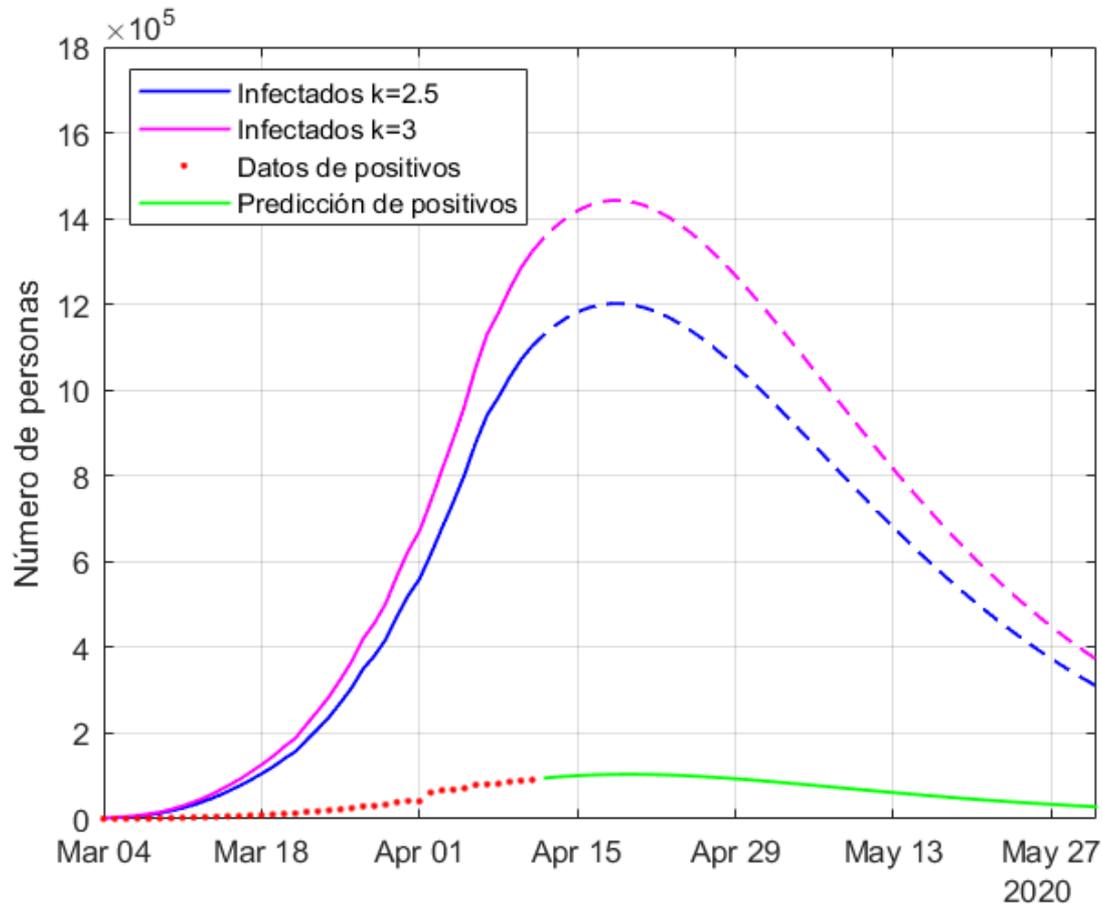
Análisis de la tasa de crecimiento en otros países. Francia



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 Francia.

En este país, tomando una ventana correspondiente a todos los datos disponibles desde el 6 de marzo, se prevé que el paso por 1 de la tasa de crecimiento se produzca entre el 16 y el 18 de abril.

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

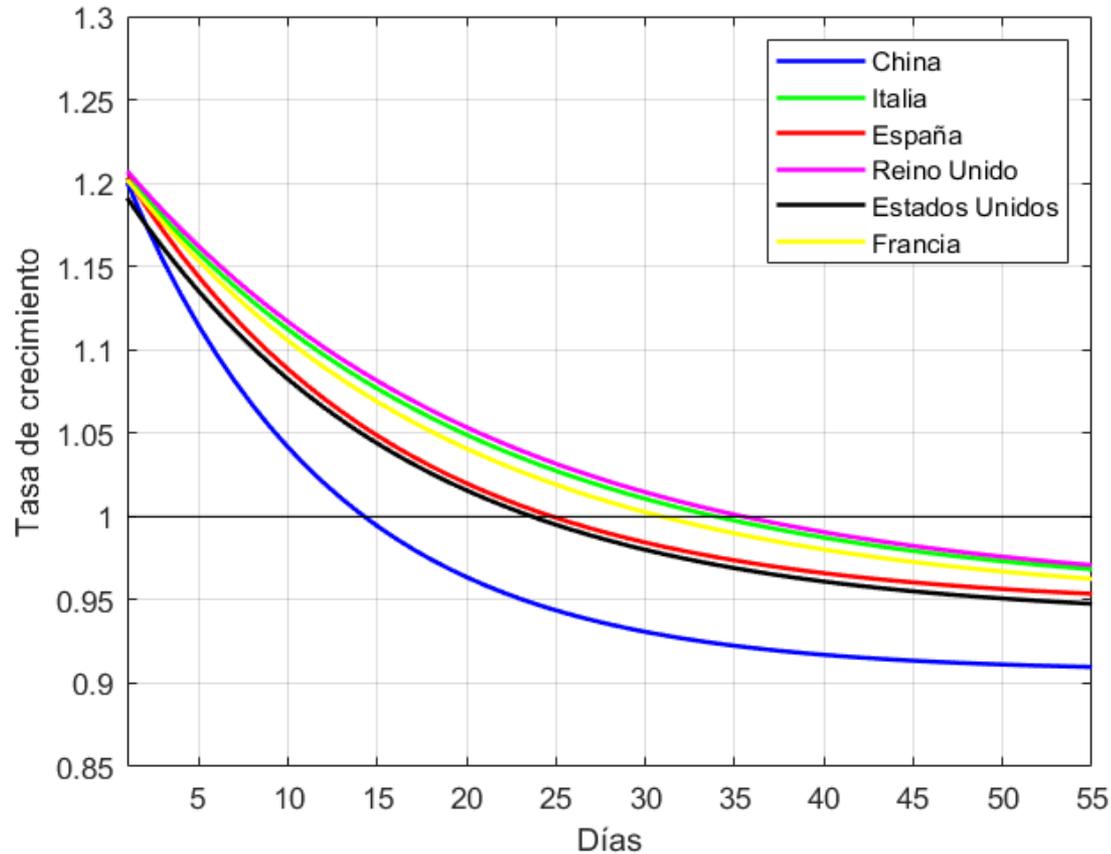


Con las mismas consideraciones tomadas en España y sus comunidades en lo que concierne al factor k , se muestra en la gráfica adjunta la predicción del número de infectados en Francia.

En este caso se comprueba que el número máximo de infectados activos estará entre 1,2 y 1,4 millones de personas.

Conjuntamente a estas curvas se ha añadido también la predicción del número de positivos activos reportados, que tendrá un pico en torno a los 110 mil.

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.

Se ha tomado como origen de tiempo el instante en que la exponencial ajustada de todos los países toma un valor representativo próximo a 1,2, es decir, el día en que el número de infectados crecía al 20% diario.

Puede verse en la gráfica que la constante de tiempo es menor en China, seguida por Estados Unidos, y España, siendo los comportamientos de Italia, Francia y Reino Unido similares, con una mayor constante de tiempo, atribuible quizá al menor grado de dureza o disciplina del confinamiento en estos países.

Así, mientras que en China se tardó 15 días en llegar al pico, desde que la epidemia crecía al 20% diario, ese lapso de tiempo es de 25 días en el caso de España, y 35 días en el caso de Italia.

Resumen de resultados para España: serie histórica

La tabla de la siguiente página recopila los principales indicadores obtenidos al aplicar nuestra metodología al caso de España, desde el primer informe publicado el 25/3 (recuérdese que cada informe utiliza los datos disponibles hasta el día anterior). Realmente, hasta el informe del día 27/3 no nos atrevimos ni se incluyó explícitamente una predicción del momento en que se produciría el pico. Entonces se preveía que el pico se alcanzaría entre el 8 y 13 de abril, dando lugar a un número de positivos en el rango de las 150 a 200 mil personas. En aquel momento no se estaban descontando aún los recuperados ni fallecidos, por lo que dicha cifra se refería a los positivos totales. A partir del lunes 30/3, empezaron a tenerse en cuenta los recuperados y fallecidos, por lo que las cifras de positivos se refieren desde entonces a los activos, lo que explica la clara reducción desde los 150-200 mil del 27/3 a los 100 mil del 30/3. También se mejoró la forma en que se sintonizan las covarianzas del filtro de Kalman. Desde el 30/3, tanto la fecha prevista para el pico, entre el 9 y 12 de abril, como la cifra máxima de positivos activos, entre 80 y 90 mil personas, se han mantenido relativamente constantes, con ligeras fluctuaciones debidas a inconsistencias en los datos recopilados por nuestra fuente (Worldometer). Respecto al número de infectados reales activos, una vez descontados los recuperados y fallecidos a partir del 27/3, el pico se ha mantenido siempre en el entorno de 1,3 a 1,4 millones de personas, asumiendo que inicialmente el número de infectados era 3 veces el número de positivos (como se ha explicado reiteradamente, si la proporción inicial de infectados fuese mayor, bastaría con aumentar los infectados proporcionalmente).

Con los datos disponibles al escribir este informe (12/4), se estima que el pico de positivos activos es de unos 88 mil, y que dicho pico, así como el de infectados activos, ha tenido lugar precisamente este Domingo de Resurrección, en línea con nuestras predicciones de hace casi dos semanas (al redactar este informe, Worldometer daba 87.231 positivos activos, mientras que la cifra del Ministerio de Sanidad era algo menor: 86.656). Las simulaciones con datos de China también han confirmado la precisión de las predicciones que hace nuestro método.

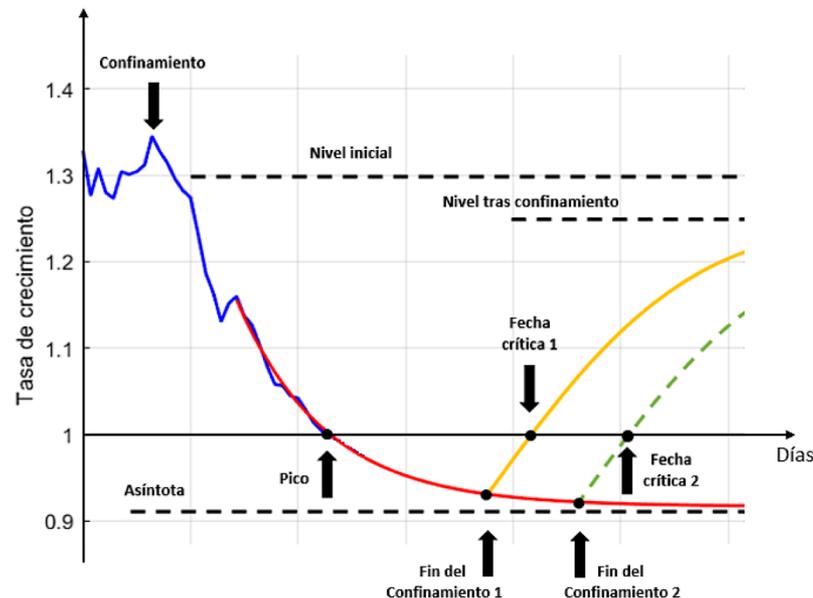
Resumen de resultados para España: serie histórica

Fecha Informe	Fecha pico (r=1)	Pico positivos activos (miles)	Pico infectados activos para k=3 (millones)	Observaciones
25/3	1/4	65	0.5	<ul style="list-style-type: none"> Se usaron datos no consolidados. Extrapolación inadecuada a la evolución prevista.
26/3	10/4	>110	>0.75	<ul style="list-style-type: none"> No se mostró expresamente el pico, solo predicción a 5 días.
27/3	(8-13)/4	150-200	1.5-2.5	<ul style="list-style-type: none"> Primer informe que muestra el pico. Extrapolación lineal con distintas ventanas.
30/3	(7-9)/4	100	1.4	<ul style="list-style-type: none"> Primer informe con auto sintonización del filtro de Kalman. Mejor filtrado de $r(n)$
31/3	(8-9)/4	75-80	1.3	<ul style="list-style-type: none"> A partir de esta fecha empiezan a tenerse en cuenta los recuperados, lo que explica el descenso.
1/4	(4-6)/4	80-90	1.3	<ul style="list-style-type: none"> La estimación del pico se adelanta por irregularidades en los datos.
3/4	(8-10)/4	95	1.5	<ul style="list-style-type: none"> A partir de aquí la previsión del pico se consolida para el 9-12/4
6/4	(9-12)/4	85	1.35	<ul style="list-style-type: none"> Se justifica teóricamente y se introduce el ajuste de $r(n)$ a una evolución exponencial (sistema de primer orden).
7/4	(9-12)/4	85	1.35	
13/4	13/4	85	1.35	<ul style="list-style-type: none"> Análisis de las posibles aplicaciones del método propuesto para el periodo de "desescalada".

Utilidad del modelo desarrollado en la fase de remisión

La siguiente figura muestra la evolución del valor estimado de $r(n)$ para España hasta el 12/4, así como el mejor ajuste posible a una exponencial decreciente (línea roja), que cruza el valor $r(n)=1$, momento del pico, precisamente el día 12/4, tendiendo asintóticamente desde el valor inicial 1,3 hasta el valor 0,93. La curva de afectados activos (tanto positivos como infectados reales), no mostrada en la figura, ha empezado a decrecer desde el día del pico, hasta que eventualmente se extinga la enfermedad, lo cual ocurriría bien entrado junio si no se alteran las condiciones de contorno actuales (con los datos de ayer, estimamos que hasta el 30 de mayo no se retornaría a la situación previa al 15/3).

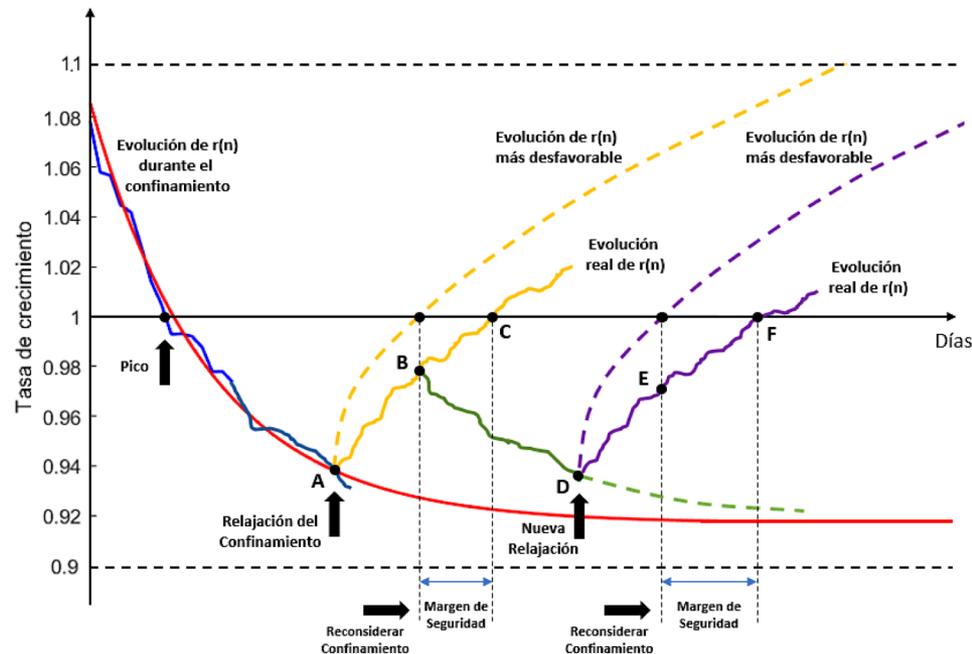
La figura muestra así mismo lo que ocurriría si un día determinado, antes de que se extinga de forma natural la enfermedad, se levantara el primer confinamiento. La constante $r(n)$ volvería a crecer de nuevo asintóticamente (línea amarilla), pero esta vez hasta un valor inferior al inicial, debido a que el número de personas susceptibles de contagiarse se ha reducido y al previsible cambio de hábitos sociales. Al crecer $r(n)$, la velocidad de reducción del número de contagiados observada hasta el momento se moderaría, e incluso la cifra de afectados volvería a rebotar el día en que $r(n)$ cruzase el umbral 1 de nuevo. La figura ilustra también claramente (línea verde a trazos) que, si el fin del confinamiento se pospone unos días, la enfermedad habrá menguado aún más, y por tanto dispondremos de más margen de tiempo hasta que de nuevo se alcance el día crítico, cuando el número de infectados volvería a crecer ($r(n)=1$). Conociendo las constantes de tiempo del fenómeno dinámico, se podría incluso estimar cuándo sería el día crítico. Esta observación nos permite finalizar nuestra serie de informes proponiendo una metodología para monitorizar diariamente el efecto de la activación o desactivación de diferentes niveles de confinamiento, como se explica a continuación.



Utilidad del modelo desarrollado en la fase de remisión

La siguiente figura muestra una vista ampliada de la banda de $r(n)$ entre 0,9 y 1,1, donde se representa la evolución ideal de $r(n)$ desde el momento del pico, de acuerdo a una curva exponencial (línea roja), en un hipotético caso donde se activaran y relajaran consecutivamente los niveles de confinamiento una vez superado el pico.

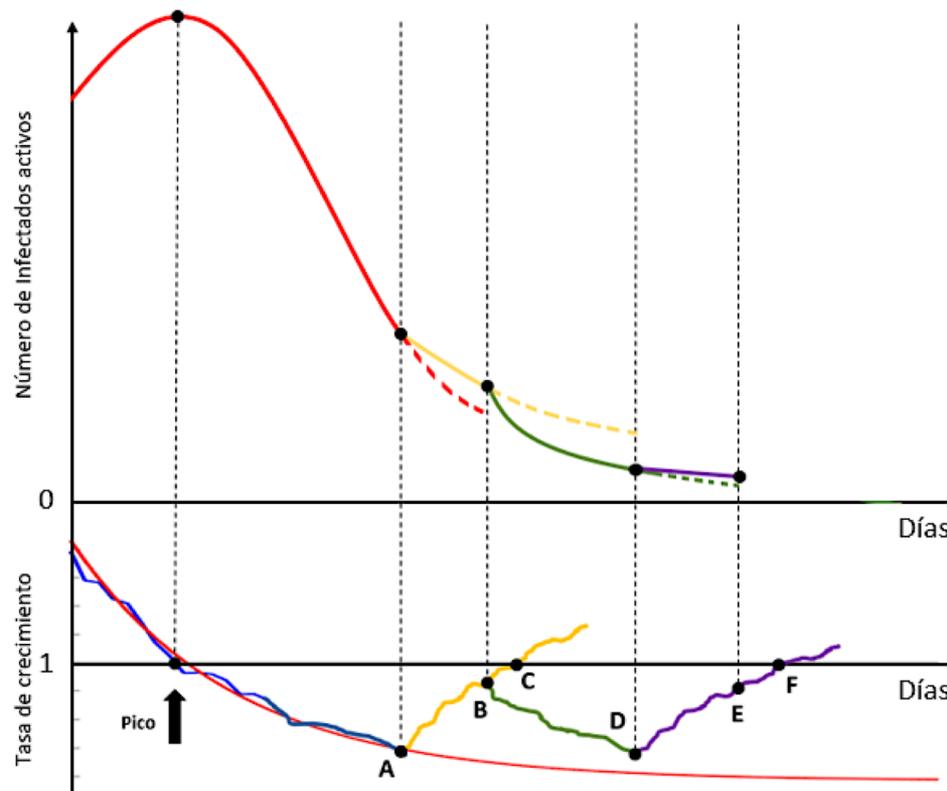
Si el día A se relajase el confinamiento que permitió superar el pico, $r(n)$ empezaría a crecer de nuevo exponencialmente, pero en el peor de los casos lo haría, como muy rápido, de acuerdo a la evolución observada en el pasado, hasta niveles iniciales ($r(n)=1,3$). En realidad, al haber menos susceptibles que al principio, y adoptarse más medidas higiénicas que antes de la epidemia, la velocidad real de crecimiento sería un poco menor, como sugieren las curvas de la figura (en amarillo a trazos la peor evolución posible, de acuerdo a lo observado en el pasado, en línea continua la evolución real prevista, algo más lenta). De ese modo, si el confinamiento se implantase de nuevo el día B, momento en que $r(n)=1$ en el peor escenario posible (es decir, vuelta a la situación de partida), existiría un margen de seguridad hasta el día C, momento en que el número de afectados comenzaría realmente a aumentar de nuevo. En otras palabras, reconsiderar la vuelta al confinamiento el día B, en lugar del C, da un colchón de seguridad que debería garantizar teóricamente que el número de afectados no se incremente de nuevo, aunque desde luego cualquier relajación del confinamiento produce una disminución de la velocidad a la que se reducen los afectados. Repitiendo el proceso cuando $r(n)$ se reduce de nuevo lo suficiente (punto D), podríamos relajar el confinamiento con cierto margen de seguridad hasta el día E (anterior al día F, que comenzaría de nuevo a progresar la epidemia). Y así sucesivamente hasta que el número de afectados se reduzca lo suficiente.



Utilidad del modelo desarrollado en la fase de remisión

Finalmente, la siguiente figura ilustra cómo evolucionaría la curva de afectados (positivos activos o infectados activos, dependiendo de la escala), de acuerdo a la evolución dinámica de $r(n)$ discutida anteriormente, desde el momento en que se supera el pico ($r(n)=1$), tras las sucesivas relajaciones y vueltas al estado de confinamiento. Se aprecia que, al levantar el primer confinamiento (día A), la epidemia decrece más lentamente (línea amarilla), pero antes de que vuelva a crecer (día B) se implanta un nuevo confinamiento hasta el día D, que logra de nuevo un descenso más acusado del número de afectados (línea verde), y vuelta a empezar.

Conociendo las constantes de tiempo de los fenómenos dinámicos involucrados, que el filtro de Kalman ha demostrado dar con suficiente precisión, podría incluso estimarse en qué días concretos (A y D) debería levantarse el confinamiento, para garantizar un mínimo margen de seguridad entre los días en que este se vuelve a implantar (B y E) y los días críticos (C y F) en que el número de afectados volvería a crecer de nuevo realmente.



Epílogo

Como ingenieros eléctricos, estamos familiarizados con modelos de sistemas dinámicos complejos, y con las técnicas de análisis de las series temporales de datos que surgen en dicho contexto, incluyendo no sólo magnitudes eléctricas, como fluctuaciones de frecuencia o sobretensiones de maniobra, sino otras magnitudes econométricas, como los precios de los mercados eléctricos o la producción renovable. Siendo conscientes de los esfuerzos realizados por colegas de otras áreas e instituciones en lo relativo al pronóstico del pico de la epidemia de Covid-19, número de personas afectadas y evolución posterior, en un alarde de ingenuidad y osadía, creímos que podríamos aportar nuestro granito de arena. Y, como suele suceder cuando un advenedizo se aventura en un terreno ignoto, pusimos sobre la mesa una metodología al parecer poco utilizada todavía en el ámbito de la epidemiología, basada en datos más que en los modelos que se derivan de las ecuaciones diferenciales al uso, lo que ahora se conoce como *data analytics*. No es momento de detallar las numerosas “trampas” que nos hemos encontrado por el camino, como las frecuentes fluctuaciones o inconsistencia de los datos, el no percatarnos antes de la importancia de sintonizar adecuadamente nuestros filtros, o lo escurridiza que resulta la simple definición de la constante R_0 de una epidemia, pero llegado este momento creemos honestamente que hemos logrado el objetivo que nos propusimos, como intentaremos plasmar en futuras publicaciones científicas a partir de ahora (esta vez no había tiempo de publicar antes).

Nuestras principales contribuciones pueden resumirse en los siguientes puntos:

- La utilización, como parámetro básico y más simple posible del modelo, de la constante $r(n)$ de la progresión geométrica, variable en el tiempo, en nuestra opinión más adecuada para caracterizar el modelo dinámico de la evolución temporal del número de infectados que el número en sí mismo.
- La aplicación de un filtro de Kalman no lineal para estimar lo mejor posible la evolución en el tiempo de $r(n)$, junto a técnicas propias de auto-sintonización del filtro, desarrolladas hace años por nosotros en otros ámbitos.
- La caracterización de la evolución de $r(n)$ como una curva exponencial, similar a la de un sistema de primer orden, tanto para curvas de datos acumulativas como para curvas “diferenciales” (sean casos activos o diarios). Esta caracterización, justificada teóricamente, ha sido crítica a la postre para asegurar una correcta predicción del momento del pico (paso por 1 de $r(n)$), y puede ser empleada, como se sugiere en este informe final, para seguir de cerca el impacto de sucesivas fases de levantamiento o reimplantación del confinamiento.

Lamentablemente, solo hemos sido capaces de atraer y retener la atención de poco más de un centenar de lectores, e ignoramos si este trabajo ha sido realmente de utilidad para alguien, pero nos sentimos gratificados por la realimentación recibida, que en ocasiones nos ha conducido a reconsiderar o justificar mejor nuestros planteamientos. Ojalá esta pesadilla termine pronto, pero creemos que no será antes del verano. Mientras tanto, sólo podremos socializar en lugares públicos, al menos durante unos meses, protegiéndonos con mascarillas. Lo que sí parece claro a estas alturas es que el futuro no será como solía.