

# Modelado y análisis de la evolución de una epidemia vírica mediante filtros de Kalman: el caso del COVID-19 en España

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: Los resultados mostrados en este documento reflejan los datos disponibles a 29 de marzo de 2020. Se incluye sólo la sección de Resultados, y se remite al lector al documento original de 24 de marzo para el resto de material (Introducción, Metodología, etc.), que se puede descargar de: <https://idus.us.es/handle/11441/94508>

## Resumen

Este documento presenta de forma sucinta los resultados actualizados obtenidos al incorporar los datos de los días **27, 28 y 29** de marzo. Los principales cambios respecto al documento del viernes 27 son: 1) se ha añadido un análisis de sensibilidad de resultados frente a valores o hipótesis iniciales, el cual ha permitido mejorar la sintonización del modelo; 2) se ha hecho una comparativa entre las previsiones de los tres últimos días, que cada vez muestran una mayor coincidencia en el pronóstico.

## 1. Resultados actualizados

### 1.1. Notación

Aunque el modelo detallado aparece en el documento original, se muestra aquí un breve resumen de la notación seguida en los resultados que se van a mostrar a continuación.

$r(n)$ : 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 incluso aunque el número de infectados no llegue nunca a saberse con certeza.

$a(n)$ : 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.

$r_F$ : ratio de variación de fallecidos.

$r_p$ : ratio de evolución de positivos.

### 1.2. Análisis de sensibilidad

La técnica de estimación utilizada requiere de unas condiciones iniciales para los parámetros que se quieren estimar y de una sintonización de las varianzas (incertidumbre), tanto del modelo adoptado como de las medidas. En esta sección se hará un análisis de sensibilidad ante variaciones en las hipótesis y condiciones iniciales, a partir del cual quedará mejor justificada la elección de los mismos.

En primer lugar, la Figura 1 muestra como varía la estimación de la tasa de crecimiento ante diferentes valores iniciales considerados, para lo cual se han ejecutado 200 simulaciones en un análisis de Monte Carlo con un rango de  $\pm 20\%$  en la estimación inicial. Puede comprobarse que el efecto de estas condiciones iniciales se

extingue con rapidez y no afecta al resultado de la estimación en los días más próximos a la actualidad. Por este motivo se ha optado por seleccionar un valor intermedio para  $r(0)$ .

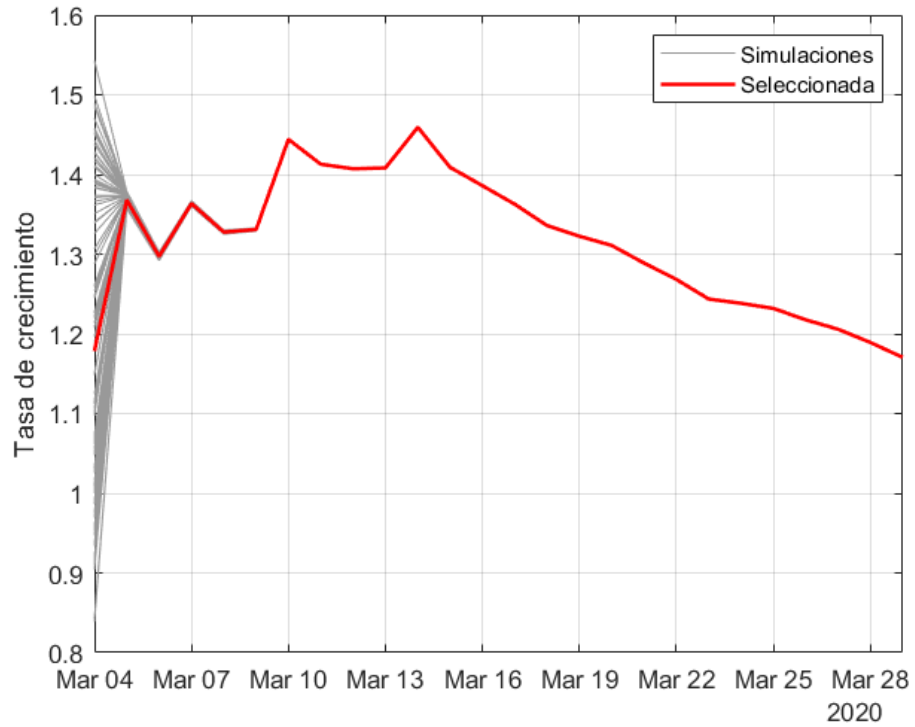


Figura 1. Sensibilidad del filtro de Kalman ante diferentes condiciones iniciales

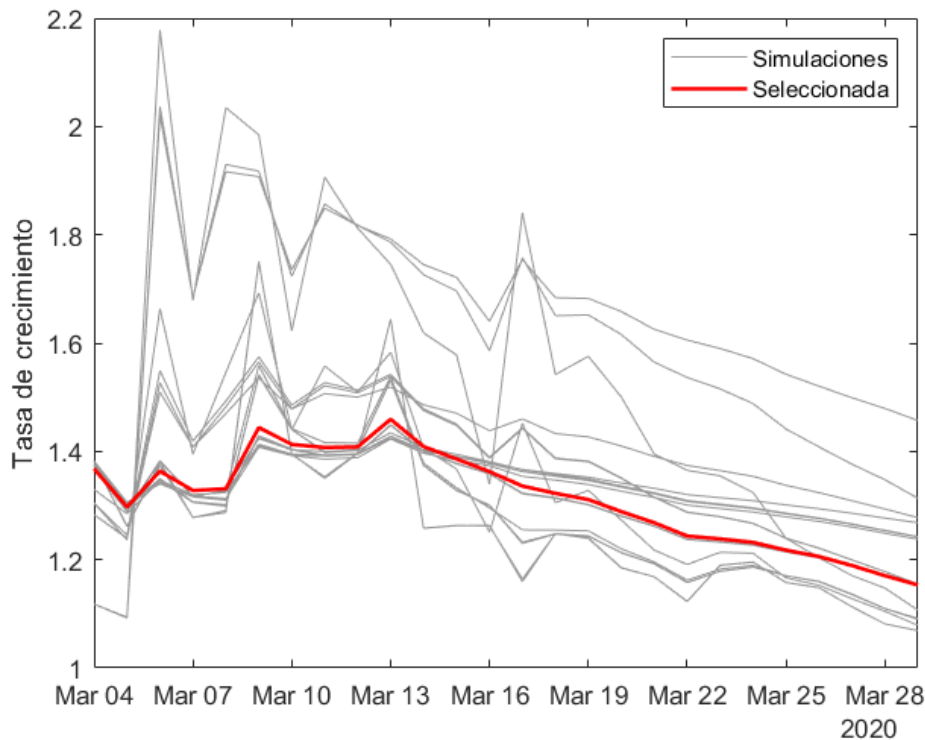


Figura 2. Sensibilidad del filtro de Kalman ante diferentes sintonizaciones

En lo que se refiere a la sintonización del filtro de Kalman, ésta se basa en ajustar los valores de las covarianzas del ruido en el proceso y en la medida. De nuevo mediante un análisis de Monte Carlo, se representa en la Figura 2 como se ve afectada la estimación de  $r(n)$ .

En este caso sí se observa una fuerte dependencia con estos parámetros, pudiendo conseguirse estimaciones de  $r(n)$  con mucha inercia, y por tanto menos fluctuaciones (filtrado intenso), pero que no son capaces de adaptarse con la suficiente rapidez a los posibles cambios de la tasa de crecimiento. También aparecen otras situaciones en las que la estimación de  $r(n)$  fluctúa en exceso, debido a lo errático de las medidas (sobre todo positivos reportados), que dificultan la extracción de información. Teniendo presentes estas consideraciones, se ha adoptado una solución de compromiso para la sintonización de covarianzas, de forma que el rizado de  $r(n)$  no sea excesivo, pero permitiendo a la vez un seguimiento de las variaciones propias de la tasa de crecimiento.

### 1.3. Estimación de la tasa de crecimiento

Una vez seleccionadas las condiciones iniciales y la sintonización del filtro de Kalman, en la Figura 3 se muestra el resultado correspondiente a la estimación de la tasa de crecimiento  $r(n)$  para España, junto con los valores de las tasas  $r_F$  y  $r_P$ , correspondientes a los fallecidos y los positivos respectivamente, estas últimas calculadas en base a datos brutos. Puede apreciarse cómo el estimador basado en filtro de Kalman es capaz de "alisar" las discontinuidades, poniendo más claramente de manifiesto la tendencia de  $r(n)$ . Por ejemplo, los datos del 25 de marzo reflejaron una aparente reducción de la tasa de fallecidos,  $r_F$ , pero ello no implica que necesariamente fuese a mantenerse en los próximos días, porque días atrás ya se produjeron reducciones bastante mayores, como puede apreciarse en la gráfica (línea roja), que luego resultaron ser un espejismo. De hecho, el día 26 dicha tasa ha vuelto a crecer, lo cual confirma que resulta arriesgado pretender sacar conclusiones en base a los datos del último día.

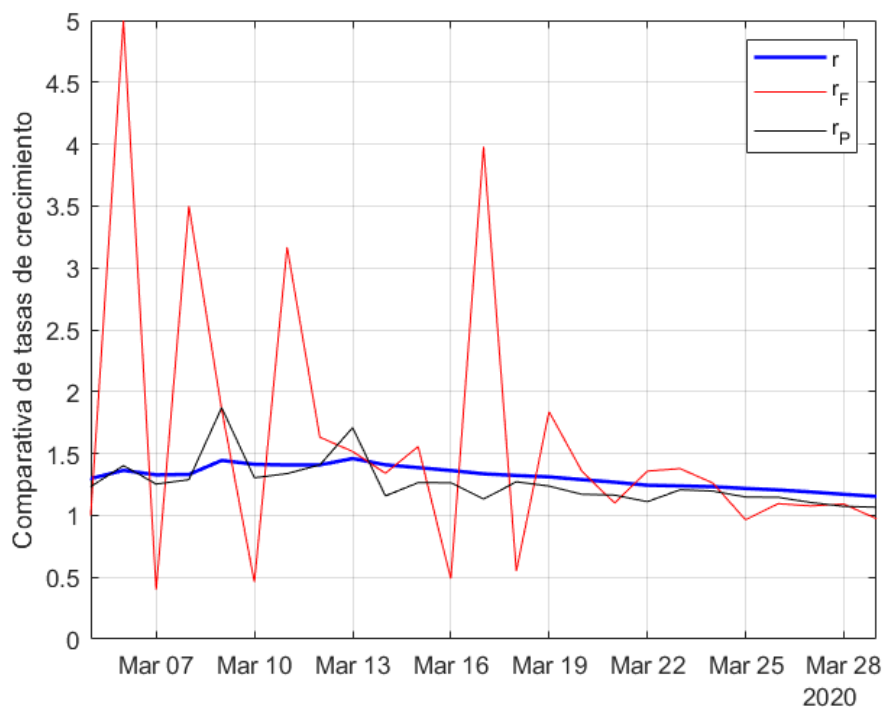


Figura 3. Comparación de las tasas de crecimiento obtenidas con datos publicados y mediante filtro de Kalman para el territorio español

En la gráfica también se aprecia claramente que la tendencia de la tasa de crecimiento estimada es ligeramente decreciente desde aproximadamente el 13 de marzo, acercándose lentamente a 1. La velocidad a la que esta disminución tiene lugar se muestra en la Figura 4, donde se observa que, tras una serie de fluctuaciones en los primeros días, coherentes con un proceso de filtrado como el propuesto, que tarda unos días en estabilizarse, a partir del 14 de marzo aproximadamente,  $r(n)$  está decreciendo a una velocidad que en los últimos días parece estabilizarse en torno a  $-0.02/\text{día}$ . Con toda probabilidad esta velocidad aumentará en los próximos días, como consecuencia del endurecimiento del confinamiento social decretado por el Gobierno desde el 30 de marzo.

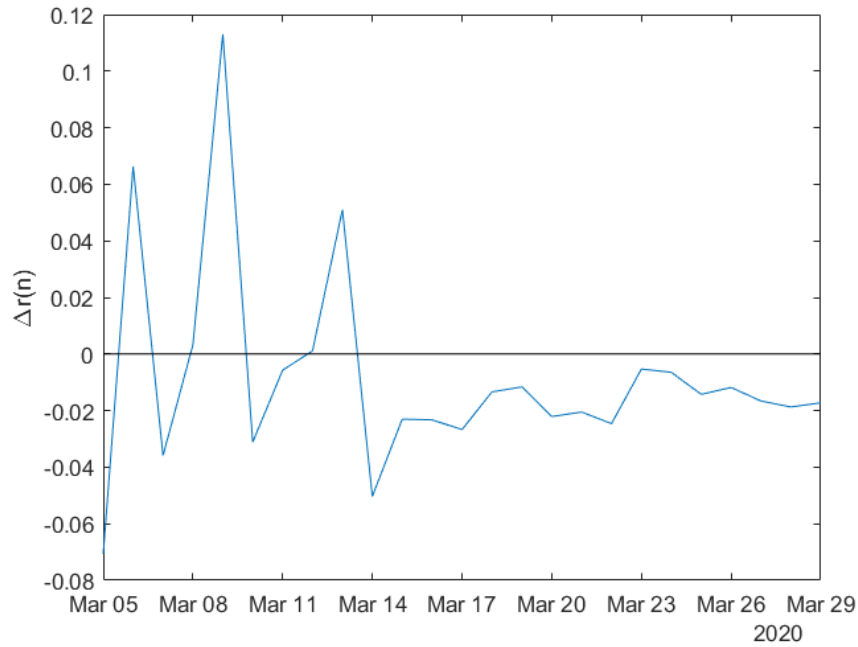


Figura 4. Tasa de variación diaria del parámetro  $r(n)$  para el territorio nacional.

Por otra parte, el mismo procedimiento se ha aplicado a los datos correspondientes a las comunidades de Madrid, Cataluña y Andalucía, mostrándose las estimaciones de sus tasas de crecimiento en la Figura 5. A diferencia de lo que ocurría en informes anteriores, en estos momentos las tres comunidades autónomas parecen ir mejorando, aunque el efecto es más acusado en Cataluña.

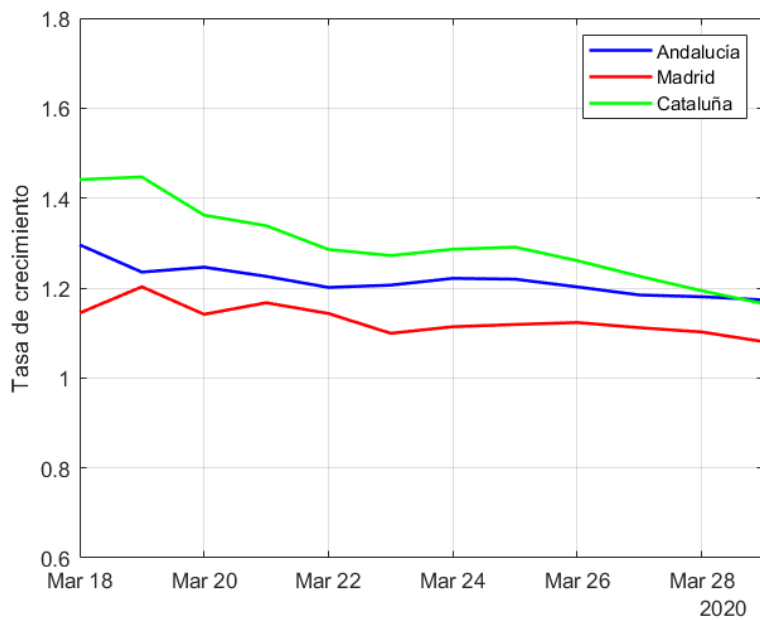


Figura 5. Tasa de crecimiento estimada para diferentes comunidades

## 1.4. Predicción

Con los resultados obtenidos en el apartado anterior, se ha realizado una predicción o pronóstico del comportamiento de la tasa de crecimiento para los próximos días en el territorio español. Dicha extrapolación a futuro se está haciendo mediante una regresión lineal, con diferentes tamaños de ventana.

La Figura 6 presenta, con los últimos datos disponibles, los valores esperados en los próximos días para la tasa de crecimiento, usando para la regresión lineal las últimas 3, 5 y 7 estimaciones de  $r(n)$ . En la representación se ha incluido un indicador de la bondad del ajuste realizado, el coeficiente  $R^2$ , cuyo valor se aproximará más a la unidad conforme mejor se ajuste la nube de puntos a una recta. Como era de esperar, el ajuste es mejor cuando sólo se utilizan los tres últimos días. A cambio, dicha predicción es más volátil que la que se hace por ejemplo con una ventana móvil de 7 días. En este punto, conviene tener en cuenta que el propio parámetro  $r(n)$  ya es el resultado de un filtrado, por lo que no puede sufrir cambios muy bruscos de un día para otro, a diferencia de los ratios basados en datos brutos.

Se aprecia que, a nivel nacional, el momento del pico ( $r(n)=1$ ), tendrá lugar entre el 6 y el 8 de Abril, en plena Semana Santa, suponiendo que la tendencia actual se mantenga, aunque podría adelantarse ligeramente debido al confinamiento más estricto a partir del lunes 30 de marzo.

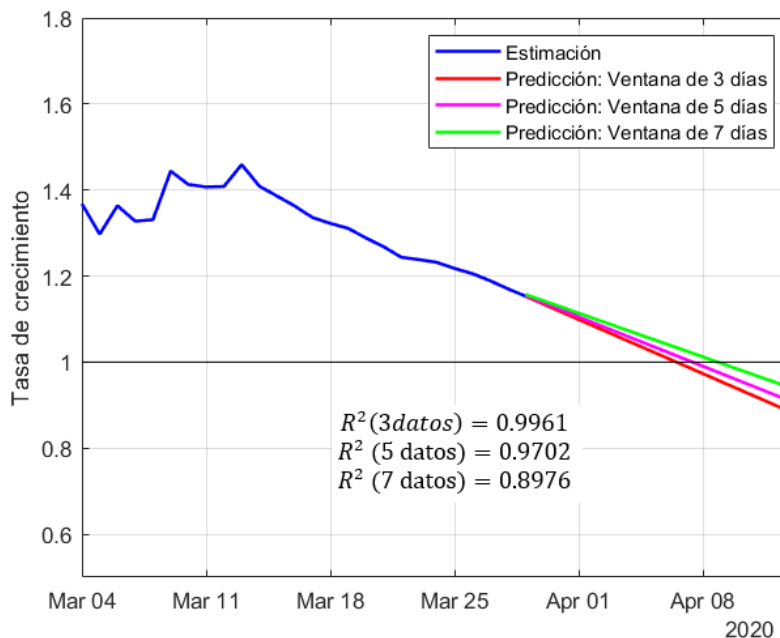


Figura 6. Estimación y predicción de la tasa de crecimiento para España con distintas ventanas de datos

Tras tres días consecutivos de análisis, puede evaluarse cómo evoluciona la tendencia de la tasa de crecimiento. En la Figura 7 se muestra el ajuste lineal con una ventana de los últimos 5 datos disponibles entre el 27 y 29 de marzo, correspondientes lógicamente al día anterior.

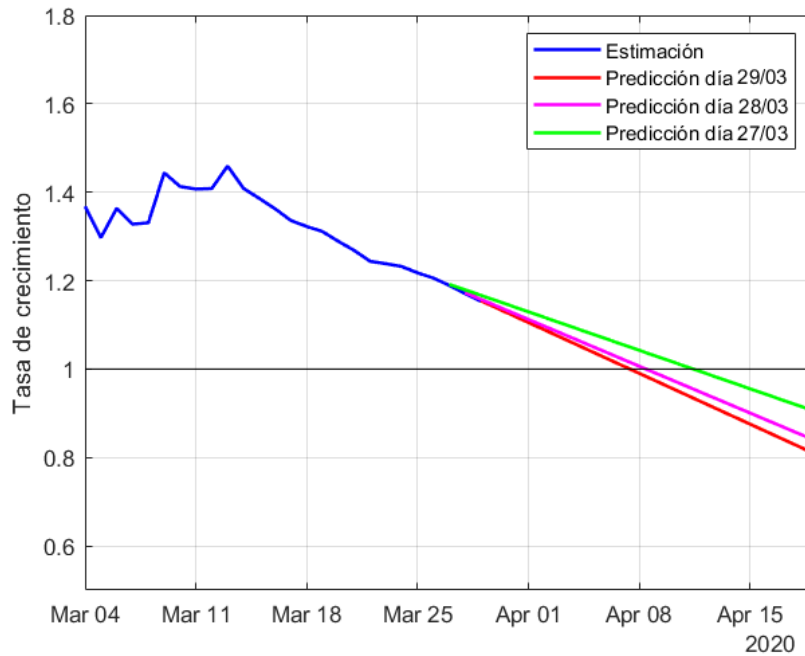


Figura 7. Evolución de la tendencia de la tasa de crecimiento en los 3 últimos días

Puede apreciarse que la tendencia de  $r(n)$  ha mejorado levemente durante el fin de semana del 27-29 de marzo, confirmándose de nuevo que  $r(n)=1$  entre el 6 y 8 de abril, si damos más importancia a los dos últimos días.

En versiones anteriores de este informe se establecía la suposición de que el número de personas infectadas por el Covid-19 en el instante inicial coincidía con el número de positivos reportados en aquel instante, estableciéndose así un factor  $k = 1$  que relaciona ambos datos. Este número, que representa una cota inferior, nos ha parecido excesivamente conservador y en esta versión se ha decidido considerar un factor más realista, entre 2,5 y 3, de acuerdo al número reproductivo básico ( $R_0$ ) reportado para este virus en los medios especializados. En la Figura 8 se muestra la banda de personas infectadas de acuerdo a dichas suposiciones ( $2,5 < k < 3$ ), en las que no se tienen en cuenta los fallecidos ni las personas ya recuperadas. En esta gráfica se ha considerado una ventana de los últimos 5 datos a fecha de 30 de marzo, para establecer la extrapolación. Puede apreciarse que, en el pico, el número de infectados estaría entre 1,1 y 1,35 millones de personas.

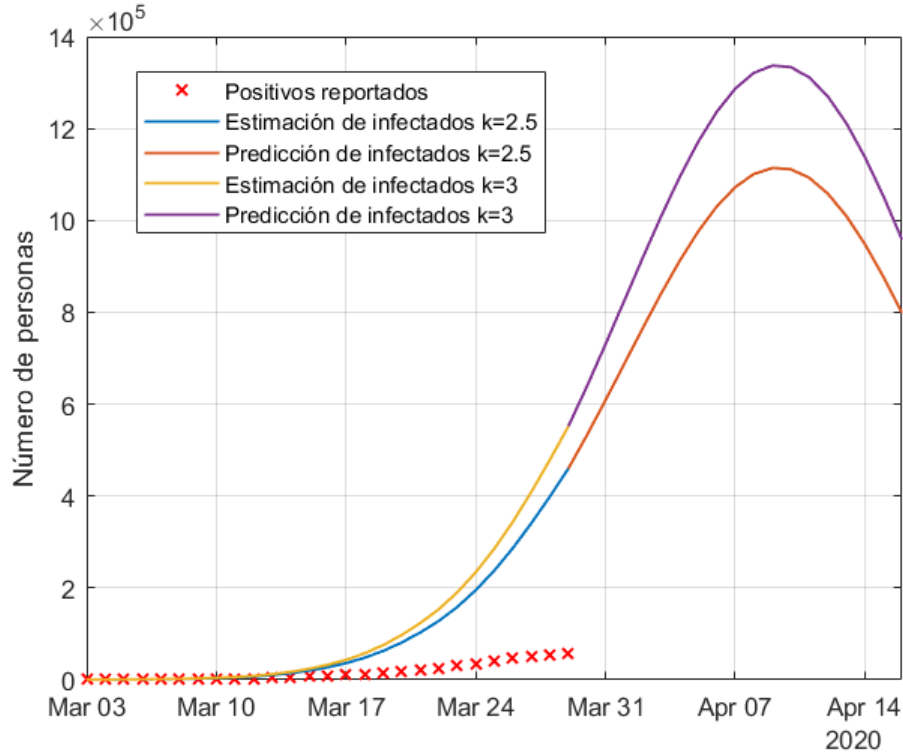


Figura 8. Estimación y predicción de infectados para diferentes valores de  $k$

En la gráfica anterior se ha incluido la predicción del número de positivos actuales o activos, descontando ya las personas recuperadas y los fallecidos.

Para diferentes ventanas de datos, se representa en la Figura 9 la predicción de positivos activos junto con el dato de fallecidos. Para la obtención de los primeros, ha sido necesario el cálculo de la tasa  $a(n)$ , cociente entre los positivos reportados y los infectados estimados, mediante filtro de Kalman, suponiendo que estos valores se pueden ajustar mediante una exponencial, que permite establecer una predicción del número de positivos reportados en los próximos días.

En el pico, se prevé que el número de positivos activos esté entre 85 y 100 mil personas.



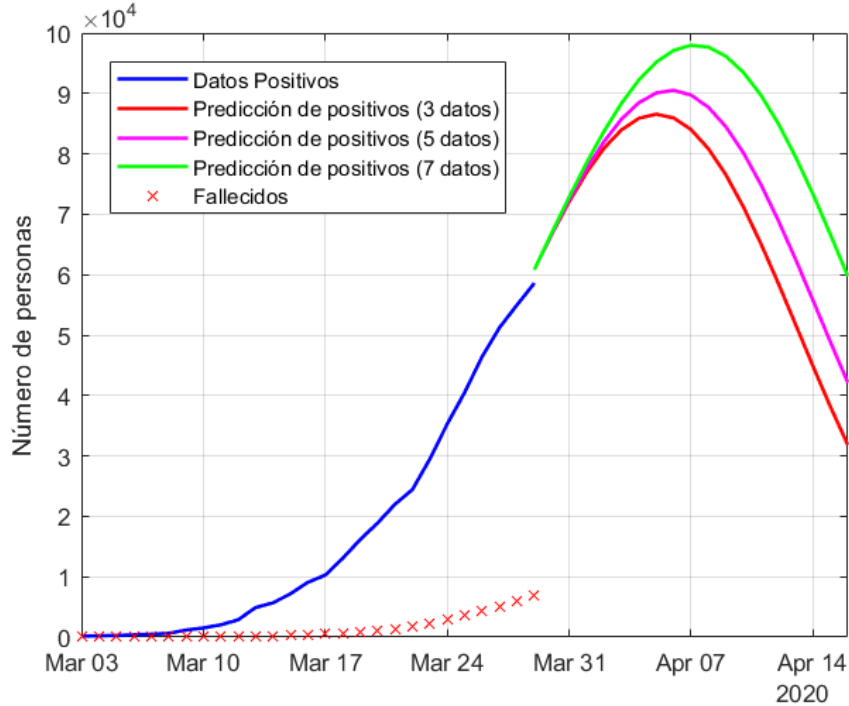


Figura 9. Predicción de positivos para diferentes ventanas de datos

La siguiente tabla recoge numéricamente los valores esperados de positivos activos en función de la ventana utilizada para hacer el pronóstico de los 5 próximos días.

Fecha	Predicción de los positivos reportados (Ventana de 3 datos)	Predicción de los positivos reportados (Ventana de 5 datos)	Predicción de los positivos reportados (Ventana de 7 datos)
30 de Marzo	66652	66736	66906
31 de Marzo	72078	72362	72853
1 de Abril	76901	77522	78516
2 de Abril	80926	82038	83738
3 de Abril	83972	85737	88357

## 2. Conclusiones

Los resultados de hoy son ligeramente mejores que los del viernes 27, y parecen ir convergiendo más nítidamente hacia la fecha del pico de la epidemia, previsto entre el 6 y 8 de abril. En el pico, se prevé que el número de positivos activos esté entre 85 y 100 mil personas, mientras que el número de infectados probablemente supere 1,2 millones de personas.