

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 25 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 y Metodología), 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 del 25 de marzo. Respecto al pronóstico sobre la evolución del parámetro $r(n)$, y por tanto del número de afectados, se han utilizado varias ventanas, aspecto fundamental que no pudo incluirse en el documento anterior por falta de tiempo.

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. Estimación de la tasa de crecimiento

En primer lugar se muestra en la Figura 1 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 *crudos*. 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 reflejan una aparente reducción de la tasa de fallecidos, r_F , pero eso no significa necesariamente que siga ocurriendo 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.

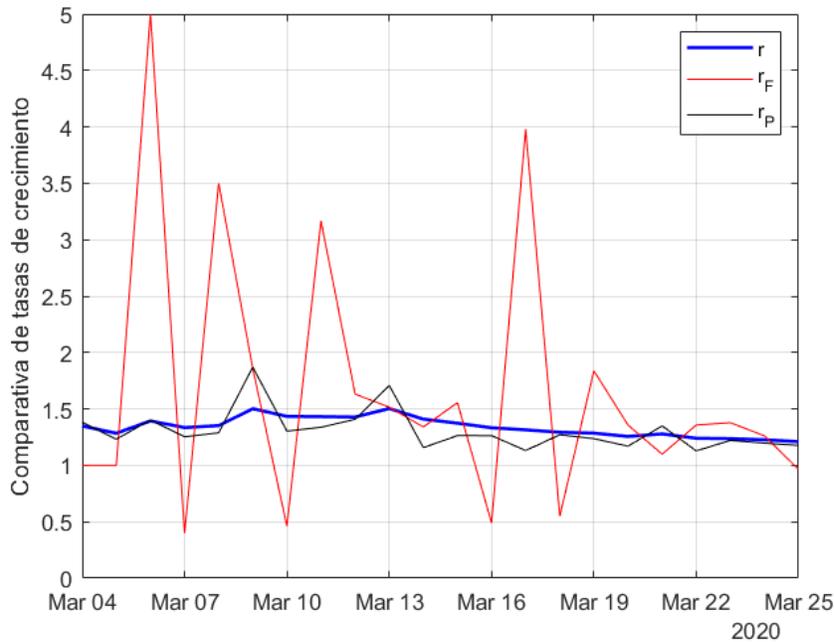


Figura 1. Comparación de las tasas de crecimiento obtenidas con datos publicados y mediante filtro de Kalman para el territorio español

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.

Por otra parte, el mismo procedimiento se ha aplicado a los datos correspondientes a las comunidades de Madrid y Cataluña, mostrándose las estimaciones de sus tasas de crecimiento en las Figuras 2 y 3 respectivamente.

Un aspecto común en ambas gráficas es que la tasa de crecimiento no sólo no se ha estabilizado aún, sino que en las últimas jornadas está sufriendo una tendencia ascendente, más acusada en Cataluña. Eso son malas noticias.

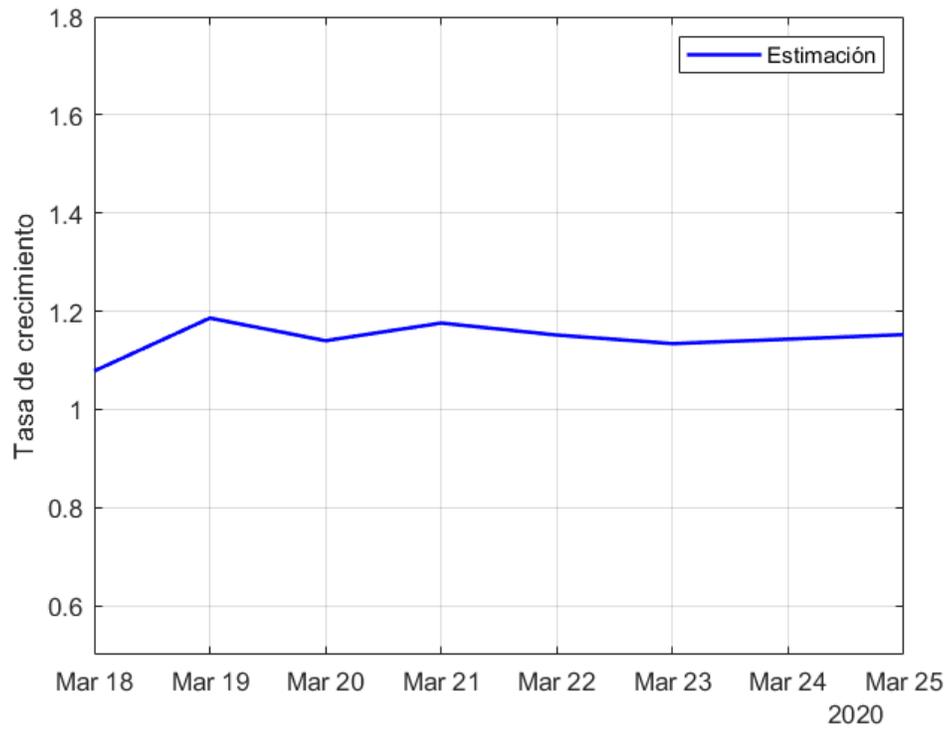


Figura 2. Tasa de crecimiento estimada para la comunidad de Madrid

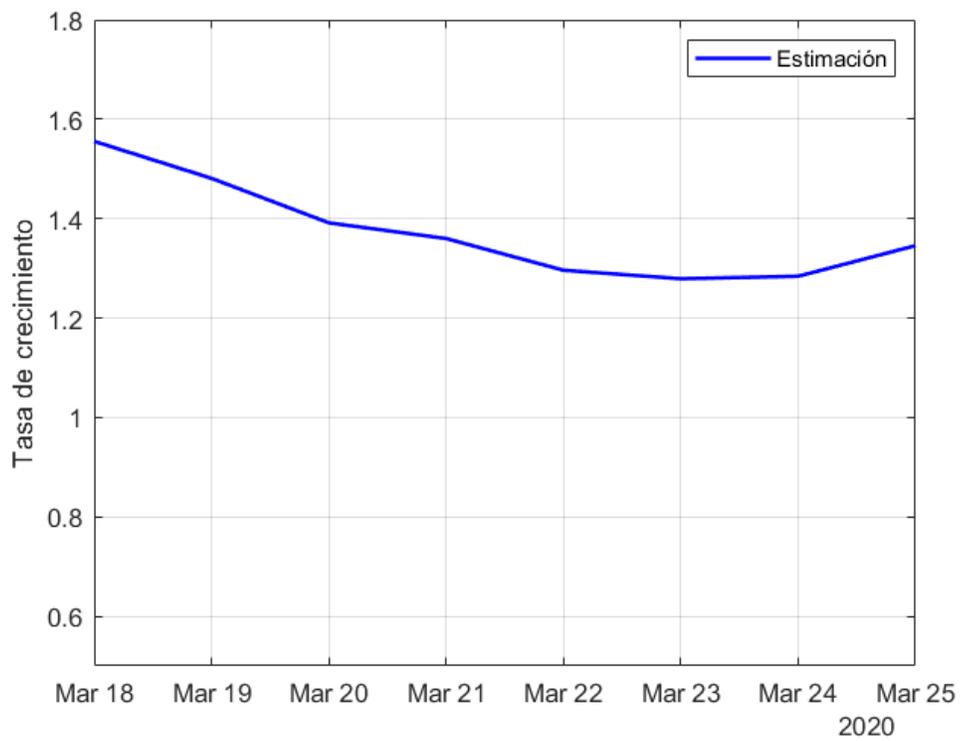


Figura 3. Tasa de crecimiento estimada para la comunidad de Cataluña

1.3. Predicción

Con los resultados obtenidos en el apartado anterior, se ha intentado realizar una predicción o pronóstico del comportamiento de la tasa de crecimiento para los próximos días en el territorio español. En el documento de ayer, esto se hizo mediante un ajuste a un polinomio parabólico sobre una ventana de los últimos datos estimados. Posteriormente vimos que el tamaño de la ventana utilizado era determinante, y que la curvatura de la parábola resultante podía ser en ambos sentidos en función de qué días abarcara la ventana. Por dicho motivo, y puesto que la evolución de $r(n)$ en el caso nacional es sensiblemente lineal en los últimos días, se ha optado en esta ocasión por hacer una regresión lineal con ventanas de diferentes tamaños.

Las Figuras 4, 5 y 6 presentan 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 $r(n)$, respectivamente. En las tres representaciones se ha incluido un medidor 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 bruscos de un día para otro, a diferencia de los ratios basados en datos brutos. Podemos adoptar una ventana de 5 días como solución de compromiso.

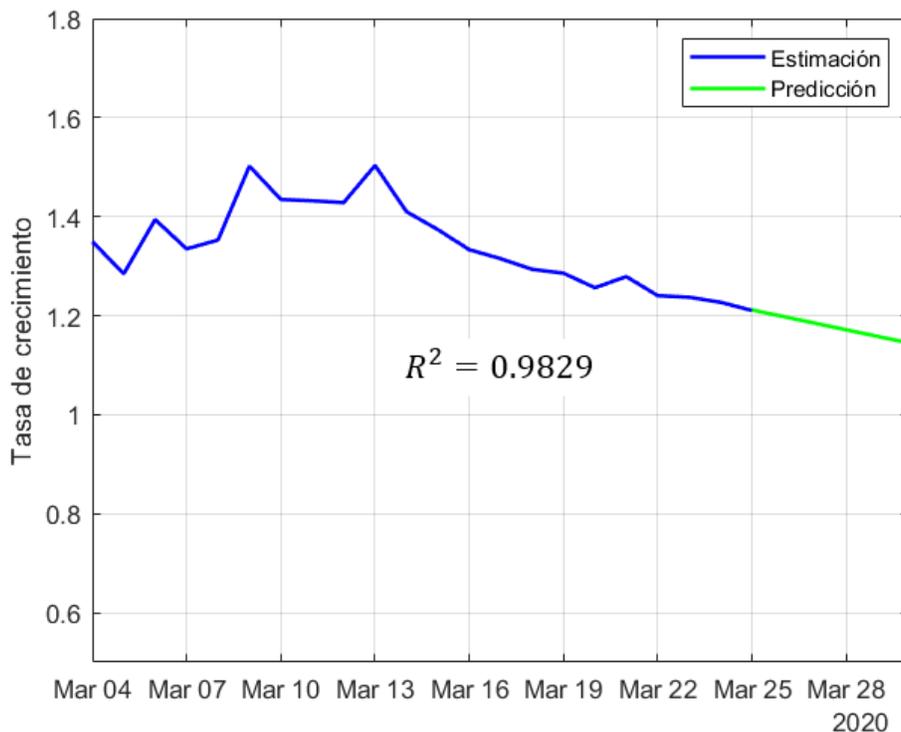


Figura 4. Estimación y predicción de la tasa de crecimiento para España con una ventana de 3 datos

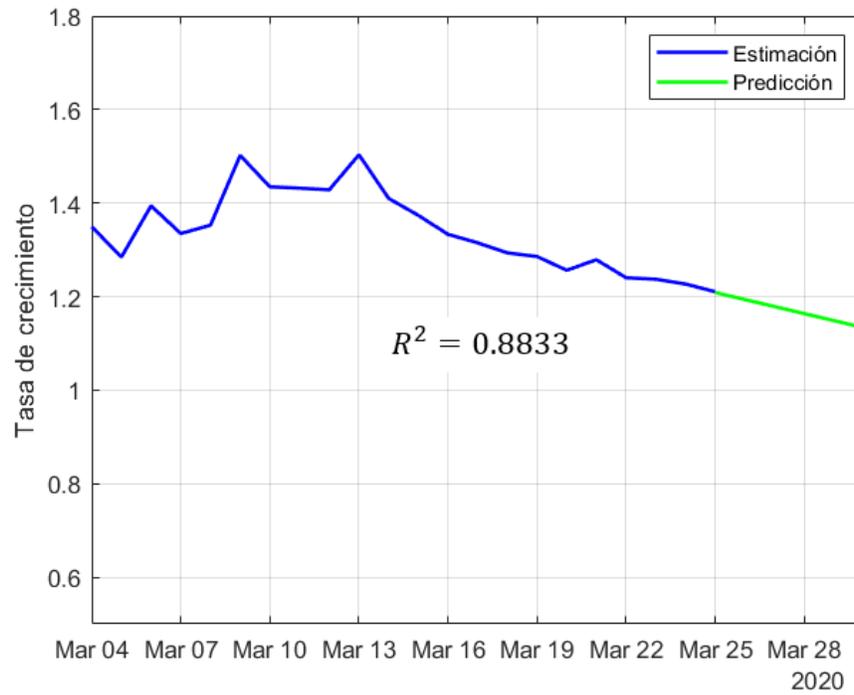


Figura 5. Estimación y predicción de la tasa de crecimiento para España con una ventana de 5 datos

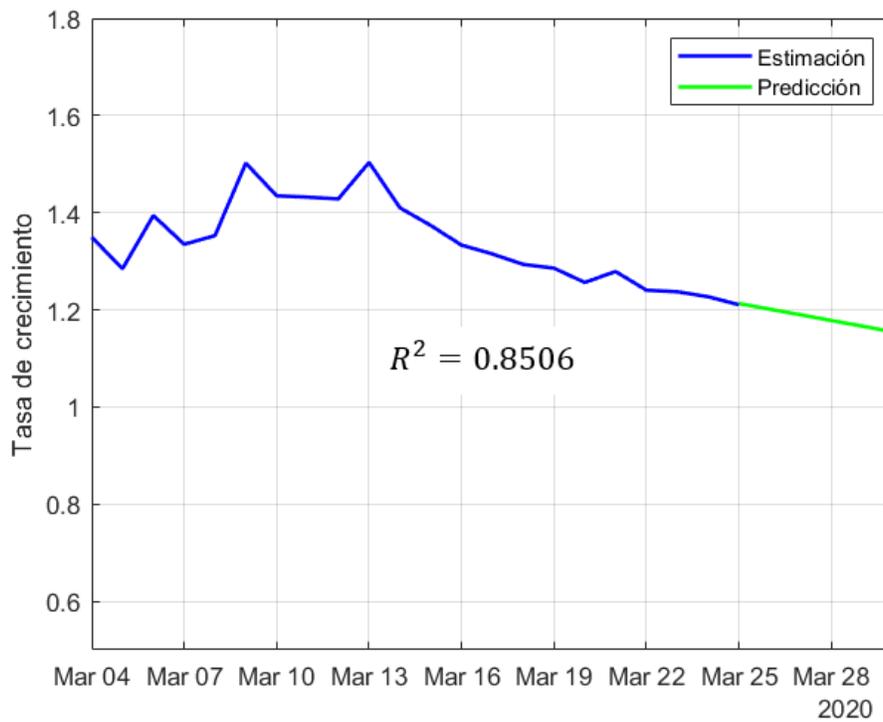


Figura 6. Estimación y predicción de la tasa de crecimiento para España con una ventana de 7 datos

Con estos valores, suponiendo, como ya se comentó en el documento anterior, un número de infectados inicial igual al de positivos reportados (que es el menor valor posible), se puede obtener la predicción de la cota inferior de los infectados reales para los próximos 5 días, tal y como se muestra en la Figura 7 para las diferentes ventanas de datos utilizadas. Se ha incluido también en la representación el número de positivos reportados cada día, pudiendo apreciarse de esta manera la diferencia entre éstos y una cota inferior muy conservadora de afectados reales.

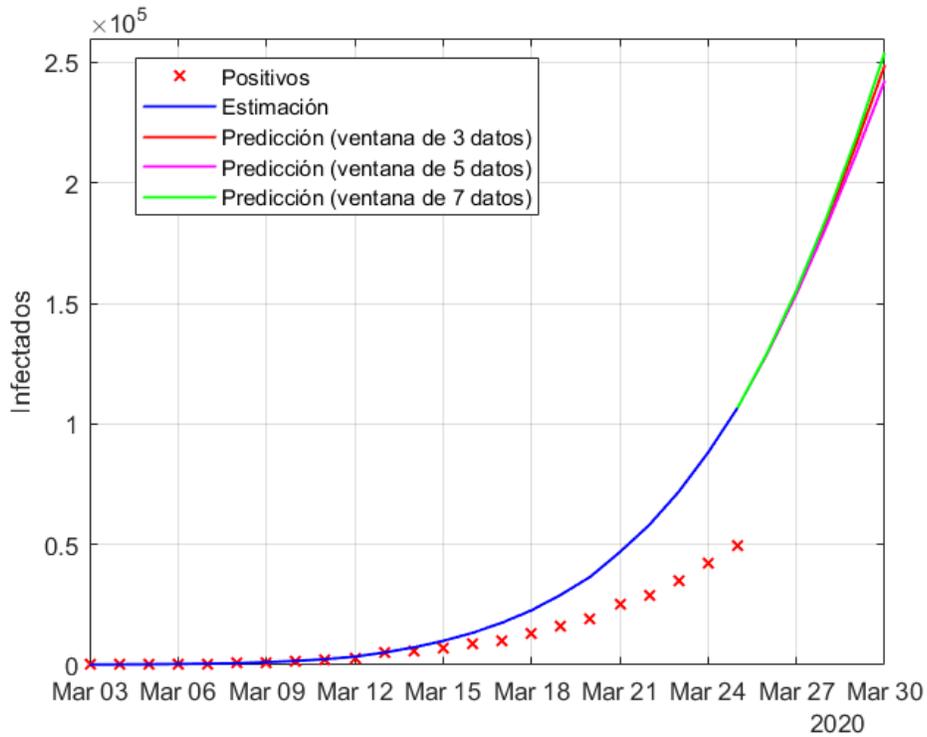


Figura 7. Estimación y predicción de infectados para diferentes ventanas de datos

Con el fin de mejorar la visibilidad de la predicción, en la Figura 8 se muestra una ampliación de la gráfica anterior.

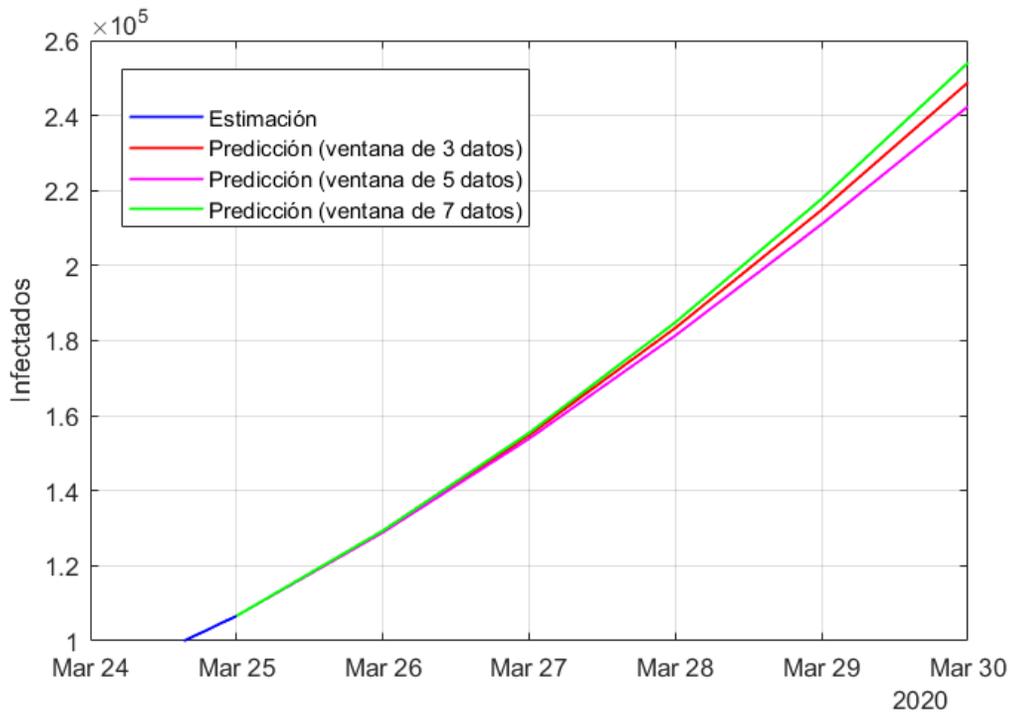


Figura 8. Ampliación de la predicción de infectados para diferentes ventanas de datos

Puede apreciarse que las tendencias son similares para las diferentes ventanas de datos consideradas, no alcanzándose el pico de infectados de la pandemia en ninguna de ellas en los próximos 5 días.

A continuación, para la tasa $a(n)$ es preciso evaluar el cociente entre los positivos reportados y los infectados estimados mediante filtro de Kalman, suponiendo que estos valores se pueden ajustar a un polinomio cuadrático, que permite establecer una predicción del número de positivos reportados en los 5 próximos días. Este resultado se muestra en la Figura 9 para las diferentes ventanas de datos, donde se incluyen tanto los datos de positivos y fallecidos como la predicción de positivos. En la Figura 10 se muestra una versión ampliada de los valores de predicción. Obsérvese que, con cualquiera de las ventanas, el pronóstico es que se superen los 100.000 positivos a finales de mes, una cifra mayor que la de ayer, en consonancia con una menor reducción prevista para $r(n)$ en función de los nuevos datos procesados hoy.

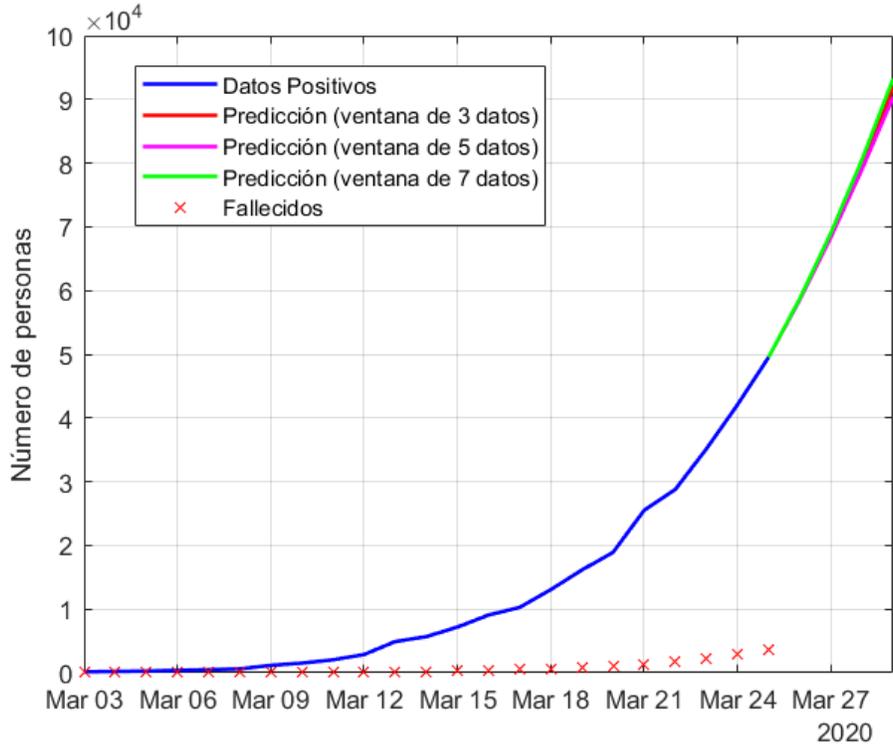


Figura 9. Predicción de positivos para diferentes ventanas de datos

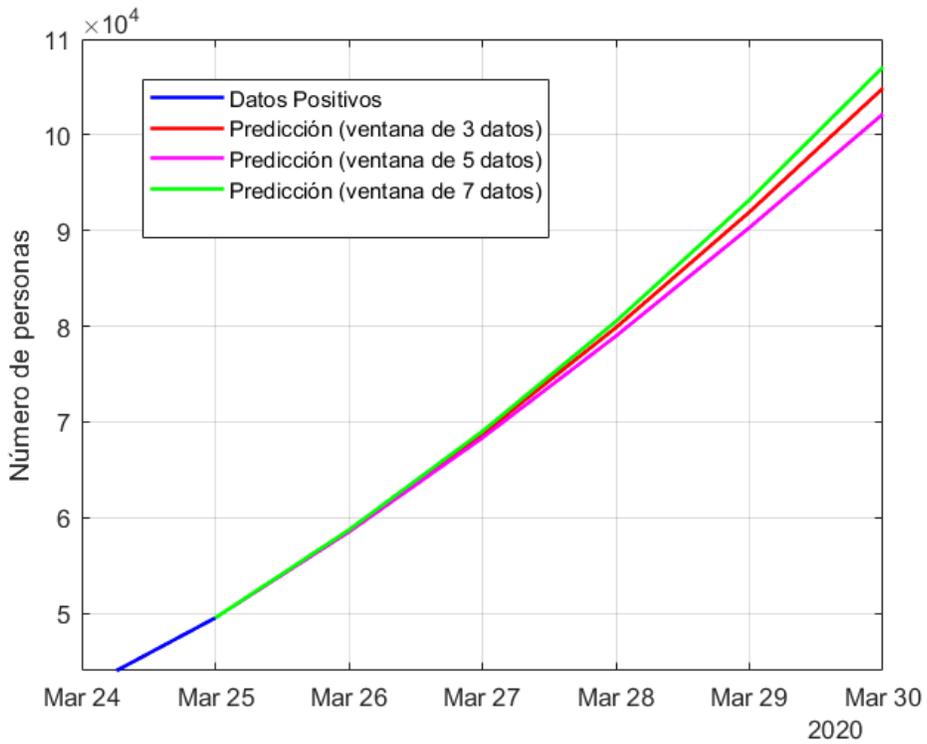


Figura 10. Ampliación de predicción de positivos para diferentes ventanas de datos

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

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)
26 de Marzo	58630	58495	58703
27 de Marzo	68735	68323	69006
28 de Marzo	79854	78966	80500
29 de Marzo	91932	90307	93199
30 de Marzo	1048880	102188	107088

El lector podrá legítimamente preguntarse por qué hay tanta discrepancia entre las predicciones realizadas por diferentes investigadores (las predicciones del pico publicadas oscilan entre dos y más de veinte días), e incluso por qué las hay entre las predicciones del mismo grupo de un día para otro. Creemos que la figura 11 permitirá entender de forma intuitiva la razón. En la parte superior se muestra la evolución de la constante $r(n)$, junto a su banda de incertidumbre, en un hipotético caso en el que su valor desciende muy rápidamente (que lamentablemente no es el caso de España). Como el cruce con la línea $r(n) = 1$ es casi perpendicular, la incertidumbre del día en que se alcanzaría el pico es comparable a la incertidumbre de $r(n)$. En cambio, en la parte inferior, donde $r(n)$ decrece muy lentamente, y se aproxima a $r(n) = 1$ tangencialmente, el momento exacto en que eso ocurre es mucho más incierto que la incertidumbre que ya de por sí trae $r(n)$. Y no sólo eso, el valor estimado es muy sensible a pequeñas variaciones de $r(n)$. Eso es justamente lo que está pasando en España, donde el ancho de la banda de incertidumbre del momento en que se alcanzará el pico es más que notable en el momento actual.

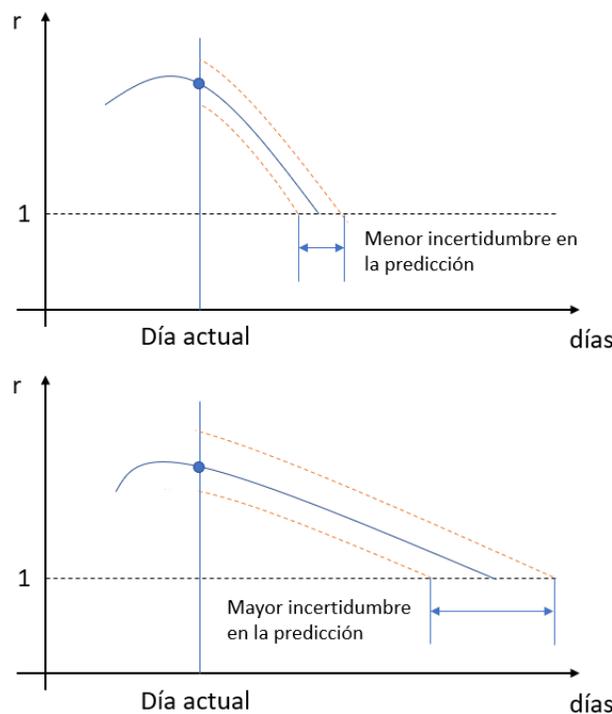


Figura 11. Ilustración esquemática de la dificultad de determinar el momento exacto del pico de afectados en el momento actual

2. Conclusiones

Los resultados de hoy son peores que los de ayer. El pico puede tardar todavía unas dos semanas en tener lugar, posiblemente debido al efecto de arrastre de comunidades como Madrid y Cataluña, que tienen mucho peso en el cómputo global. Igualmente, el número de positivos se estima mayor que ayer (posiblemente más de 100.000).