

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 26 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 del 26 de marzo. Los principales cambios respecto al documento de ayer son: 1) se ha añadido la evolución de Andalucía; 2) se ha añadido una gráfica con la velocidad de variación del parámetro $r(n)$, que resulta ilustrativa de su tendencia; 3) se ha hecho una comparativa entre las previsiones de los tres últimos días.

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.

$\alpha(n)$: 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.

r_F : ratio de variación de fallecidos.

r_P : ratio de evolución de positivos.

1.2. Estimación de la tasa de crecimiento

En la Figura 1.a 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 *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 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.

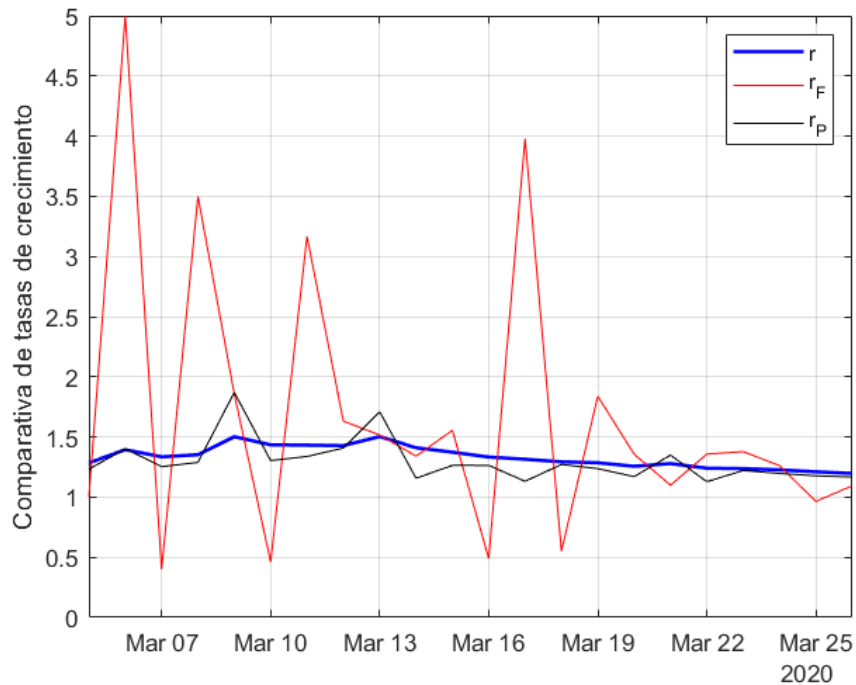


Figura 1.a. Comparación de las tasas de crecimiento obtenidas con datos publicados y mediante filtro de Kalman para el territorio español

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 1.b, donde se observa que, tras una serie de fluctuaciones en los primeros días, a partir del 14 de marzo aproximadamente, $r(n)$ está decreciendo a una velocidad que en los últimos días parece estabilizarse en torno a $-0.02/\text{día}$. Sería una magnífica noticia que esta velocidad aumentase en los próximos días.

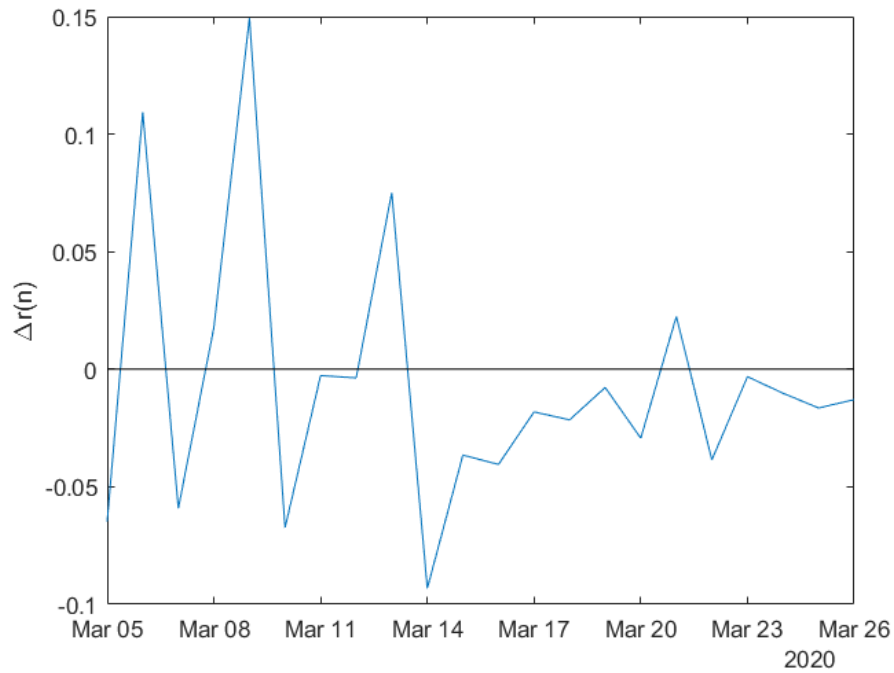


Figura 2.b. Tasa de variación diaria del parámetro $r(n)$ para el territorio nacional.

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 2. Mientras que en Andalucía puede observarse un decrecimiento lento, en la comunidad de Madrid ocurre lo contrario. Respecto a Cataluña, el último dato disponible mejora la negativa tendencia que se observaba hasta ayer.

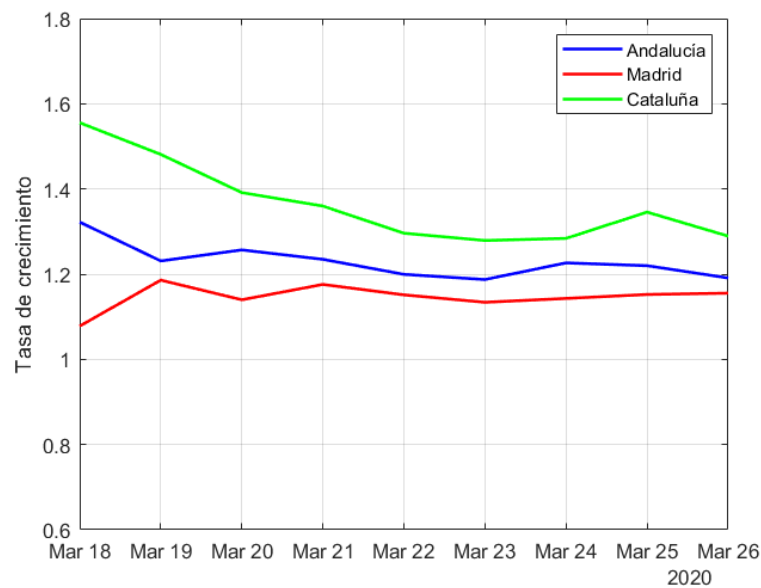


Figura 2. Tasa de crecimiento estimada para diferentes comunidades

1.3. 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. Desde ayer, dicha extrapolación a futuro se está haciendo mediante una regresión lineal (ver la discusión al final de esta sección), con diferentes tamaños de ventana.

Las Figuras 3, 4 y 5 presentan, 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 $r(n)$, respectivamente. En las tres representaciones 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 $r(n)=1$ entre el 8 y 13 de abril, dependiendo de la ventana elegida, suponiendo que la tendencia actual se mantenga, lo cual es discutible a partir del tercer día a futuro.

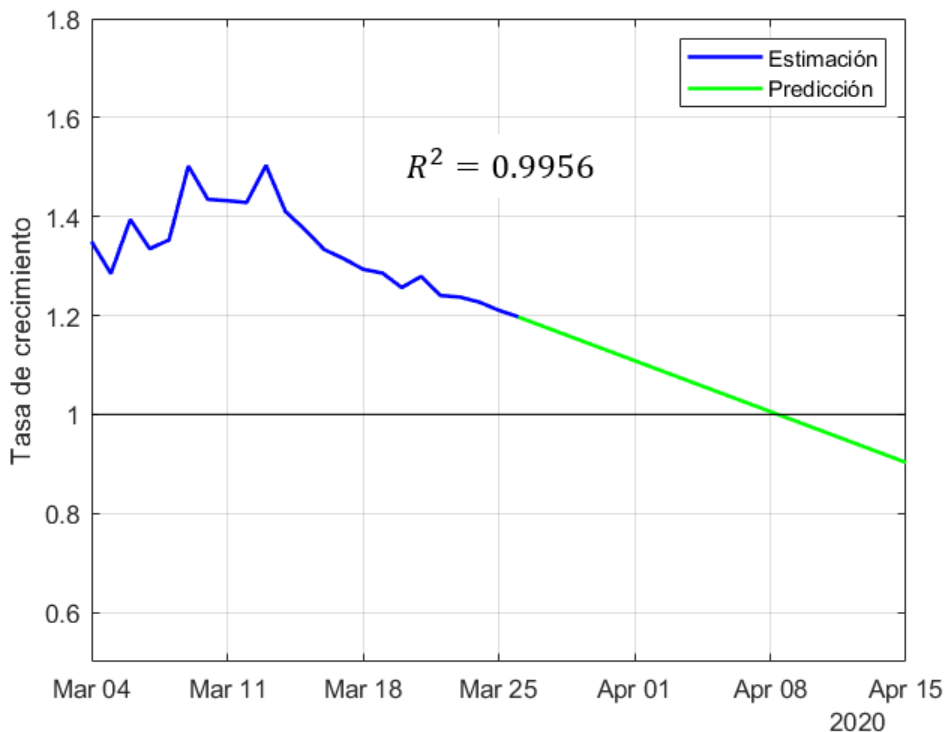


Figura 3. Estimación y predicción de la tasa de crecimiento para España con una ventana de 3 datos

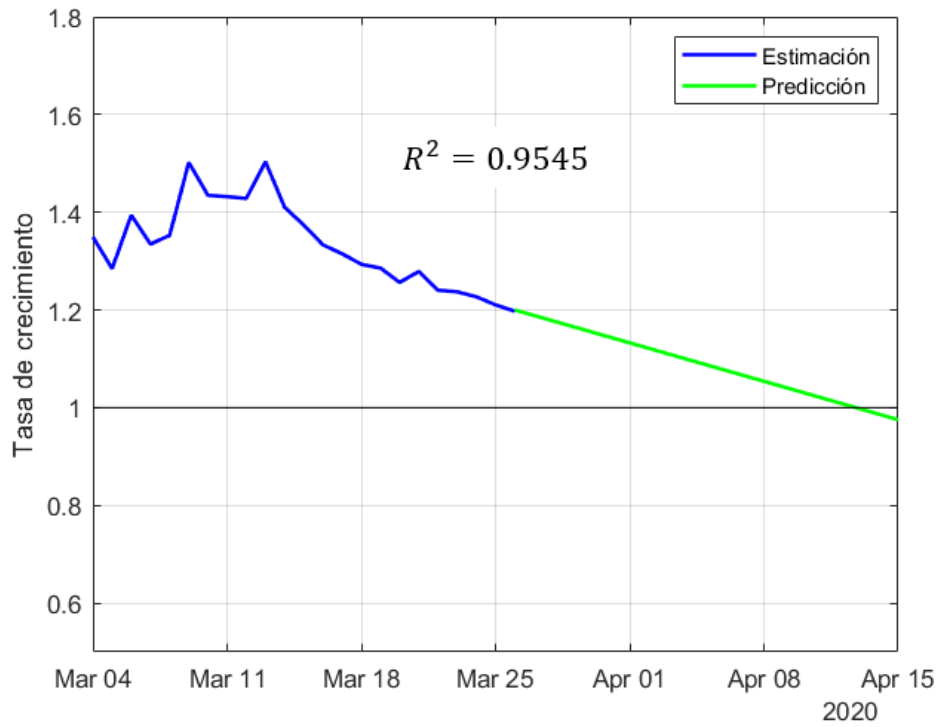


Figura 4. Estimación y predicción de la tasa de crecimiento para España con una ventana de 5 datos

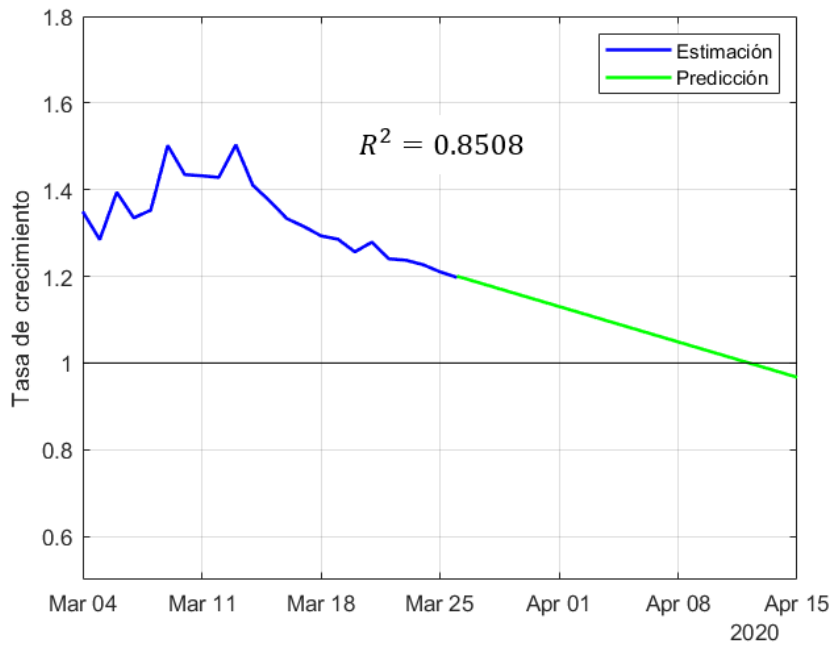


Figura 5. Estimación y predicción de la tasa de crecimiento para España con una ventana de 7 datos

Tras tres días consecutivos de análisis, puede evaluarse cómo evoluciona la tendencia de la tasa de crecimiento. En la Figura 6 se muestra el ajuste lineal con una ventana de los últimos 5 datos disponibles entre el 25 y 27 de marzo, correspondientes lógicamente al día anterior.

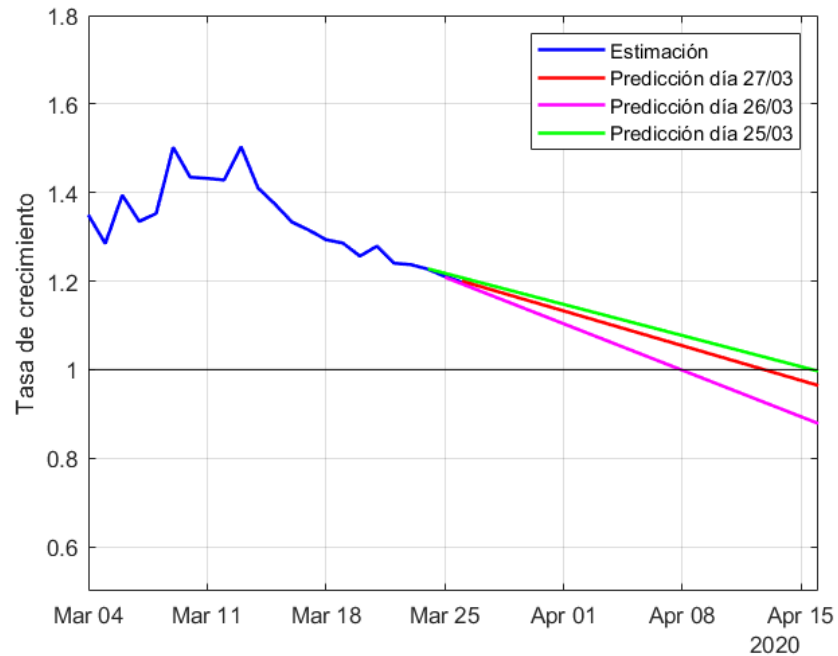


Figura 6. Evolución de la tendencia de la tasa de crecimiento en los 3 últimos días

Puede apreciarse que la tendencia de $r(n)$ ha empeorado levemente desde ayer, estando previsto que $r(n)=1$ entre el 8 y 13 de abril, si damos más importancia a los dos últimos días.

Con los valores de estimación y predicción de la tasa de crecimiento, 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 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.

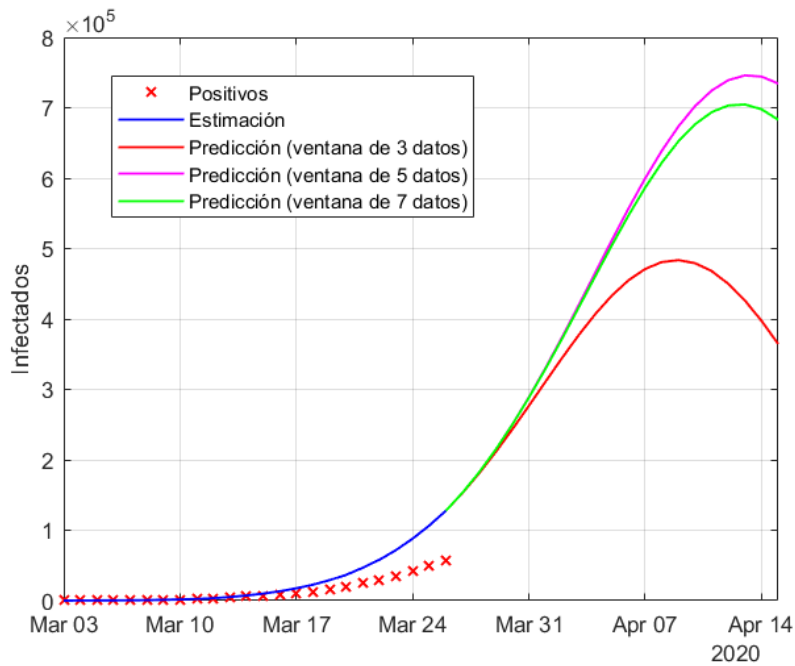


Figura 7. Estimación y predicción de infectados para diferentes ventanas de datos

Puede observarse que, con una ventana de 3 datos, el pico se alcanza antes que con las ventanas de 5 y 7 datos, que presentan resultados similares en este aspecto.

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 próximos días. Este resultado se muestra en la Figura 8 para las diferentes ventanas de datos, donde se incluyen tanto los datos de positivos y fallecidos como la predicción de positivos. Obsérvese que, con cualquiera de las ventanas, el pronóstico es que se superen los 150.000 positivos en la primera decena de abril, 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.

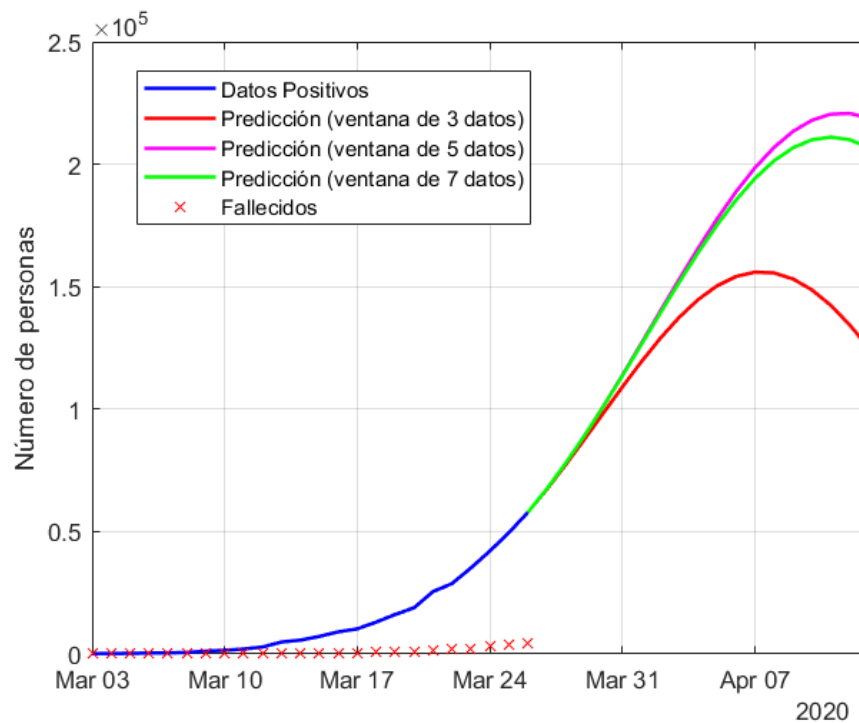


Figura 8. Predicción de positivos para diferentes ventanas de datos

La siguiente tabla recoge numéricamente los valores esperados de positivos en función de la ventana utilizada para hacer el pronóstico 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)
27 de Marzo	67142	67315	67334
28 de Marzo	77219	77847	77864
29 de Marzo	87725	89198	89178
30 de Marzo	98431	101256	101151
31 de Marzo	109071	113873	113618

Finalmente, nos parece pertinente hacer una breve discusión sobre el impacto de hacer un pronóstico basado en una regresión lineal (ayer y hoy) o en un ajuste polinómico de orden superior, tal como una parábola (anteayer). La figura 9 muestra un caso hipotético, en el que durante tres días consecutivos la evolución del parámetro en cuestión es una línea recta perfecta, pero al cuarto día el valor estimado se aparta sólo ligeramente por encima (curva superior) o por debajo (curva inferior). Puede apreciarse que el ajuste lineal (azul) apenas se ve afectado por la presencia del cuarto valor estimado, mientras que el ajuste cuadrático (rojo) cambia su curvatura completamente, conduciendo a valores totalmente diferentes a partir del sexto día. Ello quiere decir que, salvo que haya alguna razón o ley física, que lleve a pensar que el parámetro en cuestión va a evolucionar cuadráticamente, es preferible utilizar un ajuste lineal. En este caso, la evolución de $r(n)$ tiene un marcado carácter lineal, como indica la bondad del ajuste en las gráficas anteriores.

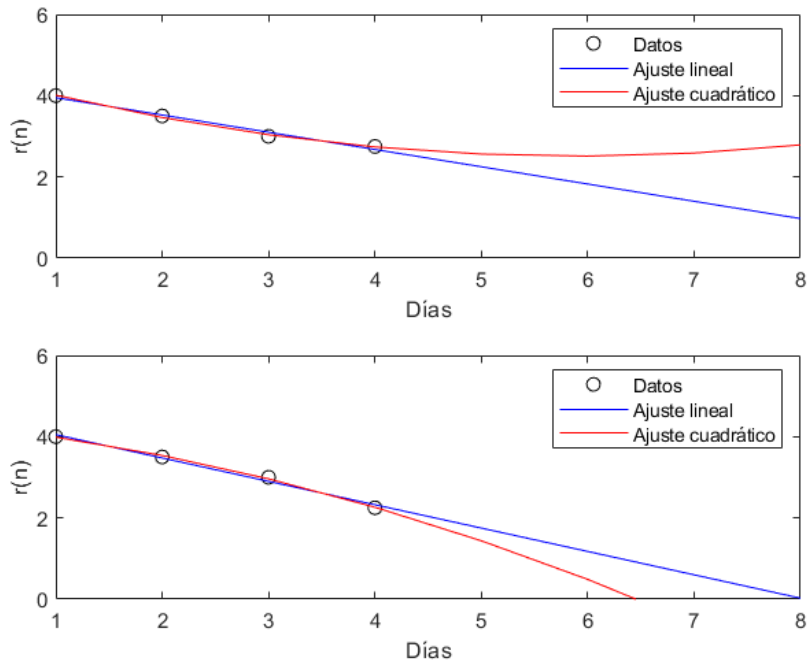


Figura 9. Ilustración esquemática de la volatilidad asociada a un ajuste cuadrático (parabólico) frente a un ajuste lineal (recta de regresión)

2. Conclusiones

Los resultados de hoy son similares o ligeramente peores que los de ayer, confirmándose en todo caso la tendencia decreciente de $r(n)$. Se estima que el pico de la epidemia puede tener lugar entre el 8 y 13 de abril, y que el número de positivos supere los 150.000.