

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 31 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: En el informe de hoy, además de actualizar los resultados con los últimos datos disponibles, se ha incluido una comparación de la evolución de $r(n)$ en España, China e Italia.

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

Una vez seleccionadas las condiciones iniciales y la sintonización del filtro de Kalman, en la Figura 1 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 volvió a crecer, lo cual confirma que resulta arriesgado pretender sacar conclusiones en base a los datos del último día.



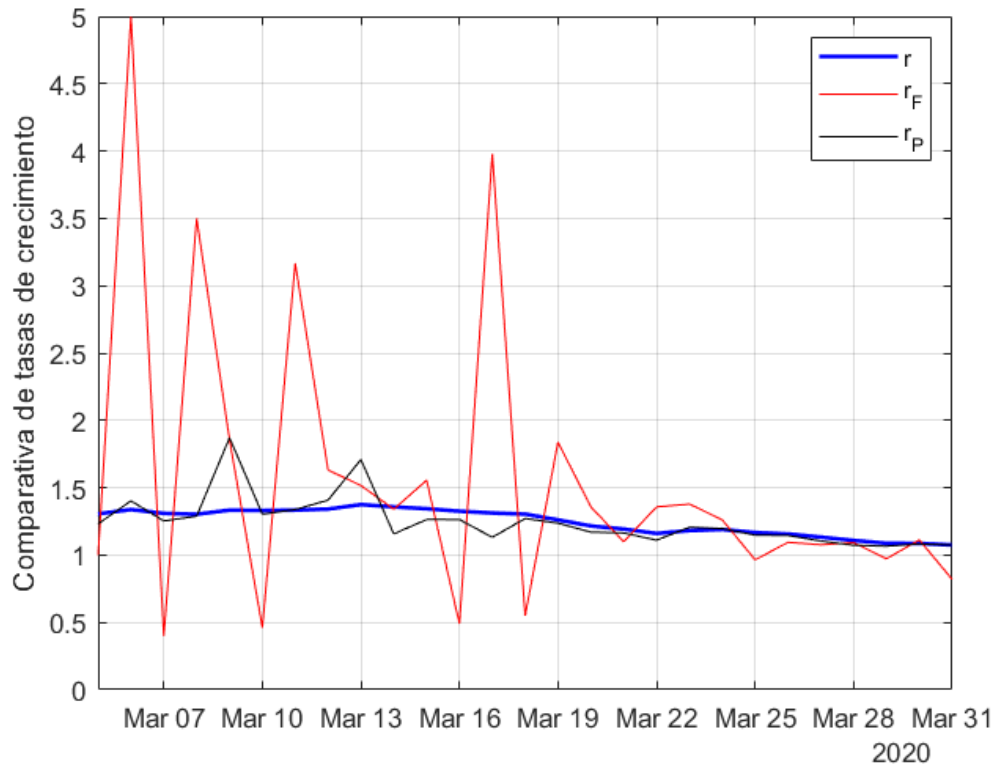


Figura 1. Comparación de las tasas de crecimiento obtenidas con datos publicados y mediante filtro de Kalman para el territorio español

En la gráfica también se aprecia claramente que la tendencia de la tasa de crecimiento estimada es ligeramente decreciente desde aproximadamente el 13-14 de marzo, acercándose lentamente a 1. La velocidad a la que esta disminución tiene lugar se muestra en la Figura 2, 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 de forma sostenida, a una velocidad que a su vez parece ir decreciendo en los últimos días, estando actualmente en torno a $-0.015/\text{día}$. Con toda probabilidad, esta desaceleración volverá a aumentar en los próximos días, como consecuencia del endurecimiento del confinamiento social decretado por el Gobierno desde el 30 de marzo.

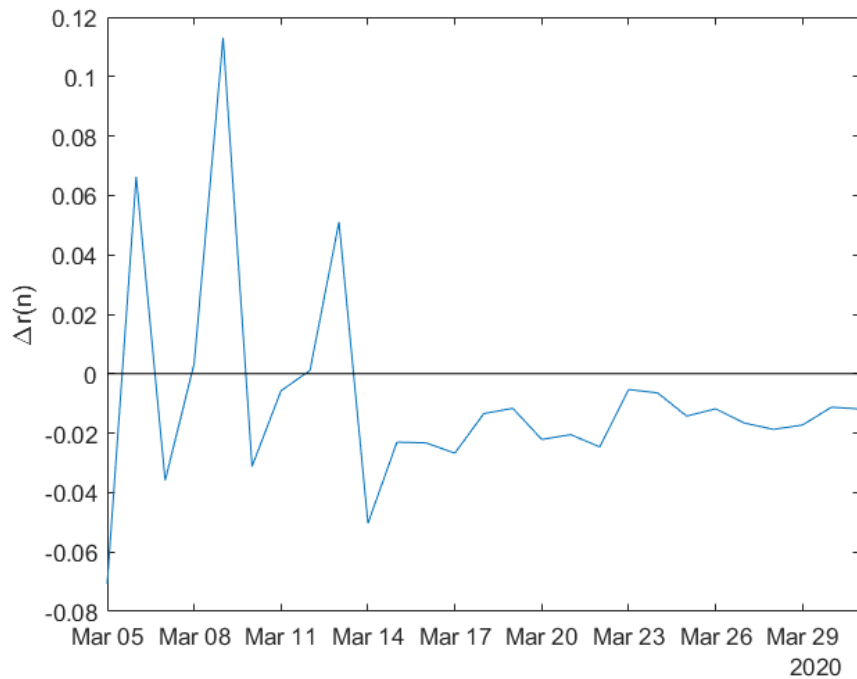


Figura 2. Tasa de variación diaria del parámetro $r(n)$ para el territorio nacional.

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 3. A diferencia de lo que ocurría en informes anteriores, en estos momentos las tres comunidades autónomas parecen ir mejorando (muy débilmente en Madrid, que es un dato preocupante por su gran peso en el cómputo nacional), aunque la desaceleración es más acusada en Cataluña.

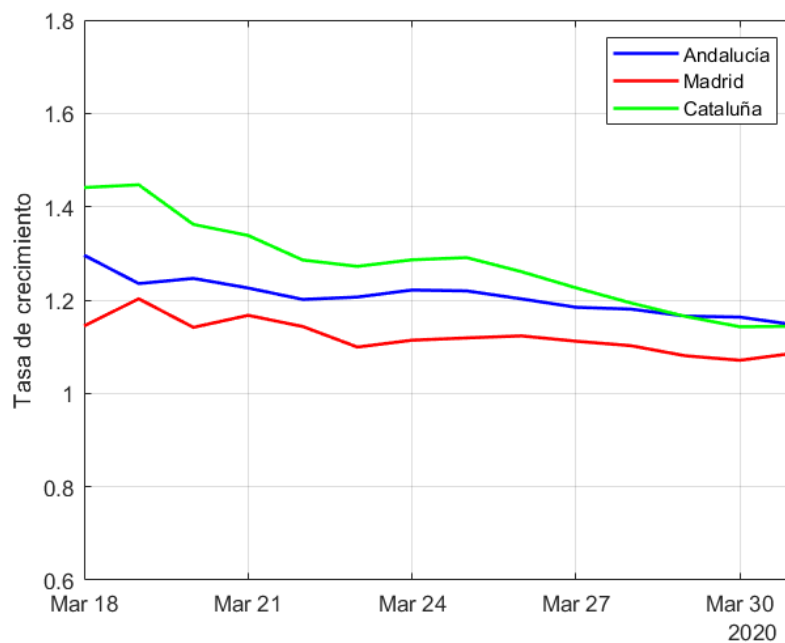


Figura 3. Tasa de crecimiento estimada para diferentes comunidades

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. Dicha extrapolación a futuro se está haciendo mediante una regresión lineal, con diferentes tamaños de ventana.

La Figura 4 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 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. A diferencia de los días anteriores, el ajuste en esta ocasión es peor cuando sólo se utilizan los tres últimos días, lo cual indica que ayer hubo una perturbación en el modo en que se registran los datos, o un cambio de tendencia a peor (esto último poco probable). Ello confirma en todo caso que la predicción basada en los tres últimos días 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 5 y 7 de abril (un poco antes de lo que se preveía en el informe de ayer), suponiendo que la tendencia actual se mantenga, si bien la predicción basada en los tres últimos días sugiere que será después (alrededor del 12). Realmente, si las nuevas medidas de confinamiento tomadas desde el 30 de marzo no surtiesen efecto, podría darse la indeseable situación de que la evolución de $r(n)$ se aplanase aún más, quedando su valor oscilando en las inmediaciones de 1, pero sin disminuir luego sensiblemente. En ese caso, más que un pico nos encontraríamos con una meseta.

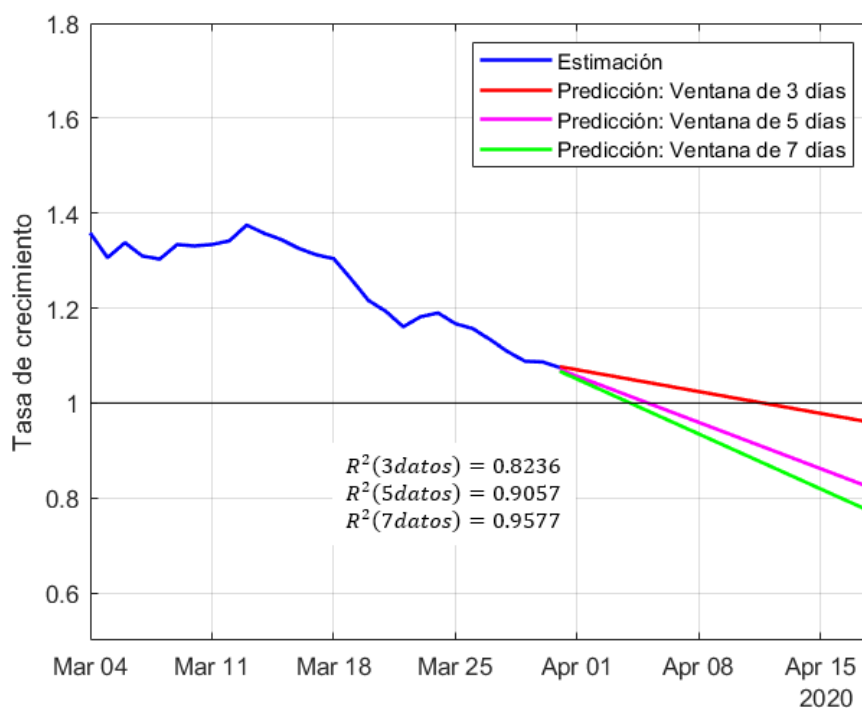


Figura 4. Estimación y predicción de la tasa de crecimiento para España con distintas ventanas de datos

Tras tres días consecutivos de análisis, puede evaluarse cómo evoluciona la tendencia de la tasa de crecimiento. En la Figura 5 se muestra el ajuste lineal con una ventana de los últimos 5 datos disponibles entre el 29 de marzo y el 1 de abril, correspondientes lógicamente al día anterior.

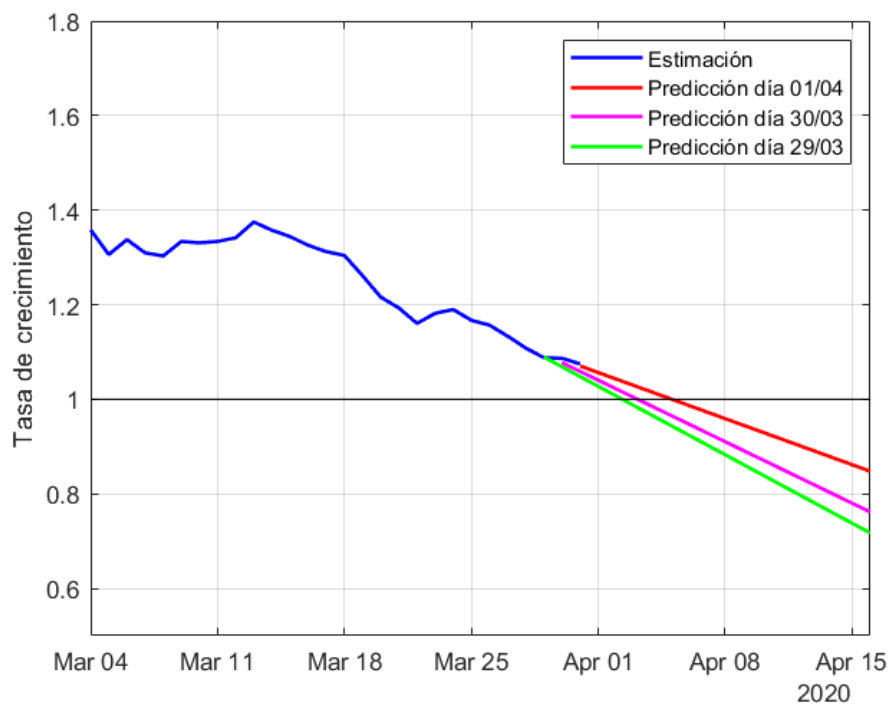


Figura 5. Evolución de la tendencia de la tasa de crecimiento en los 3 últimos días

Puede apreciarse que la tendencia de $r(n)$ ha mejorado levemente durante los últimos días, previéndose en estos momentos que el pico ocurrirá entre el 3 y 6 de abril, en contra de lo que sugerían los datos de ayer.

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 6 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 1,30 millones de personas.

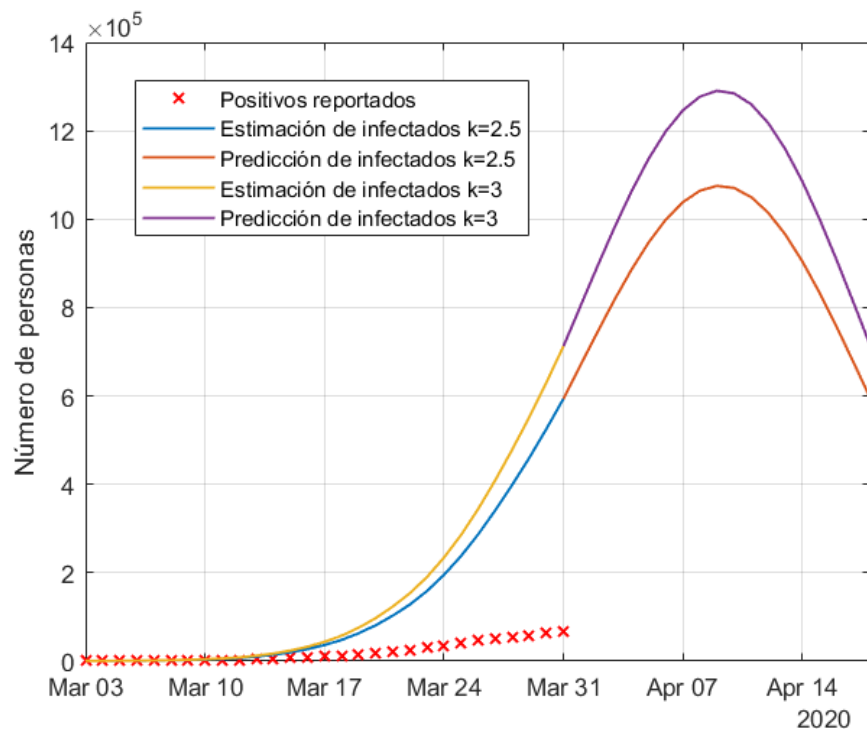


Figura 6. Estimación y predicción de infectados para diferentes valores de k

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 7 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 80 y 90 mil personas.

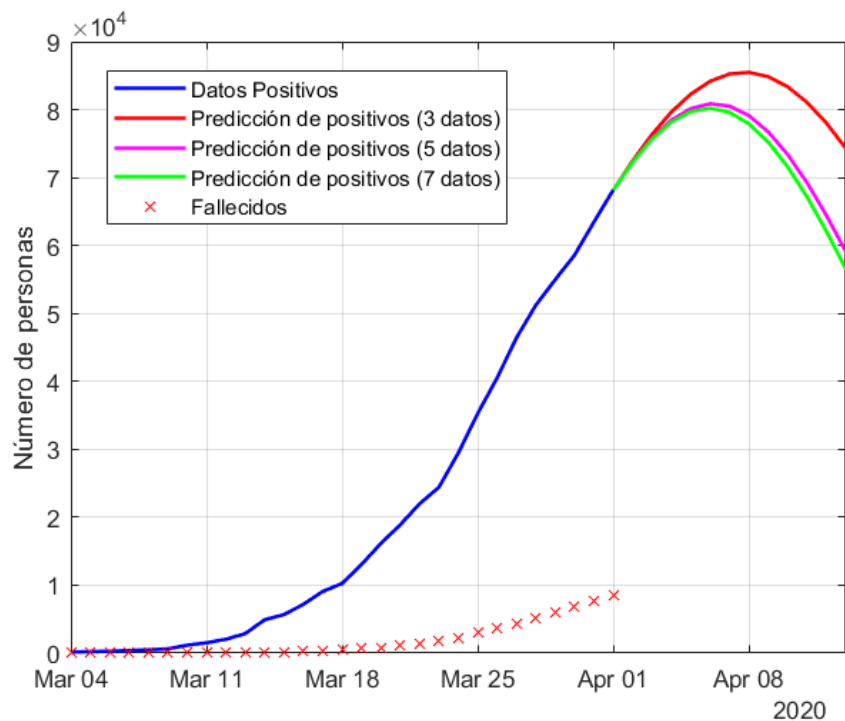


Figura 7. Predicción de positivos para diferentes ventanas de datos

La siguiente tabla recoge numéricamente los valores esperados de positivos activos (ojo, menor que el de positivos acumulados) 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)
1 de Abril	72490	72327	72282
2 de Abril	76362	75804	75673
3 de Abril	79663	78455	78191
4 de Abril	82297	80170	79727
5 de Abril	84183	80873	80208

1.4. Comparativa con otros países

En esta sección, y tras analizar los datos disponibles de España, se ha realizado una comparación entre la situación que presenta nuestro país, desde que se estableciera el confinamiento (14 de marzo) con otros países en situación similar. En concreto, se ha optado por representar la evolución de la tasa de crecimiento de España junto con la de Italia y China, tal y como aparece en la Figura 8.

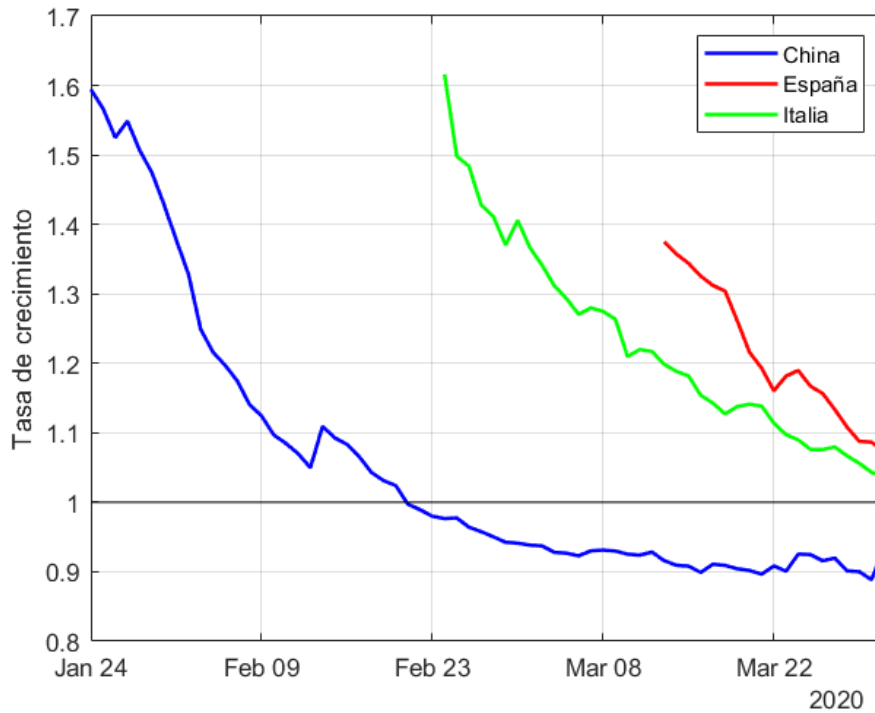


Figura 8. Comparación de las tasas de crecimiento de España, Italia y China tras el confinamiento

Tomando como referencia el caso de China, que empezó a sufrir los efectos de esta pandemia antes que el resto de países del mundo, puede verse que el comportamiento de la tasa $r(n)$ es asintótico, encontrándose el valor de la asíntota bastante por debajo de 1, afortunadamente. A simple vista, también se aprecia que el efecto del confinamiento en España es más pronunciado que en Italia, pero menos que en China.

2. Conclusiones

Los resultados de hoy son similares a los que se presentaron ayer, aunque ligeramente más optimistas, sugiriendo como fecha probable del pico de la epidemia entre el 4 y 6 de abril. En el pico, se prevé que el número de positivos activos esté entre 80 y 85 mil personas, mientras que el número de infectados probablemente supere 1,3 millones de personas.