

## Trabajo de investigación:

# Modelado y análisis de la evolución del COVID-19 en España

Grupo de Investigación TEP-196 (Sistemas de Energía Eléctrica)

### Autores:

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 informativa para los medios

Esta nota trata de aclarar el objetivo, alcance y resultados hasta la fecha del proyecto "Modelado y análisis de la evolución de la epidemia de COVID-19 en España", obviando en lo posible los detalles matemáticos, de forma que pueda ser comprendido por cualquier persona. Para más detalles, se remite al lector a los informes que se van actualizando periódicamente en el repositorio idUS: <https://idus.us.es/handle/11441/94508>

*Nota: se permite la reproducción parcial o total de este texto, citando el nombre de los autores y la filiación.*

### Objetivos

Los efectos sociales y económicos de la pandemia del Covid-19 están siendo tan devastadores que todos los agentes involucrados (gobiernos, organismos internacionales, instituciones, empresas e individuos) tienen el máximo interés en saber cómo evolucionará en el tiempo el número de afectados, con vistas por un lado a constatar los efectos beneficiosos del aislamiento social, y por otro a planificar los ya de por sí saturados recursos sanitarios, así como tomar las medidas que palién en la medida de lo posible dicho impacto.

El objetivo principal del trabajo es predecir la evolución de la progresión geométrica que caracteriza el número de infectados reales, y en consecuencia en qué momento se alcanzará el pico de la epidemia en el caso concreto de España.

### Metodología

Los científicos, ingenieros, economistas, sociólogos, etc. disponen de numerosas herramientas matemáticas y estadísticas para el tratamiento y filtrado de series temporales, con vistas a extraer información útil de los datos disponibles, por definición inciertos, tales como tendencias, patrones, periodicidad, valores medios o esperados, varianzas, etc. Existen básicamente dos categorías de modelos para procesar la información: 1) Los que intentan caracterizar la realidad "física" que genera dichos datos. En el caso de una epidemia vírica, estos modelos consideran por ejemplo qué fracción de personas están en el trabajo, en la enseñanza o en la calle, cuánto tarda en manifestar síntomas una persona contagiada y cuánto tiempo dura el periodo de contagio, etc. 2) Modelos que intentan determinar parámetros o variables explicativas desde un punto de vista puramente matemático (caja negra), sin entrar en cuáles son las causas o interacciones que explican los datos resultantes.

Nuestro modelo pertenece a la segunda categoría, y no prejuzga cuál será exactamente la forma o curva de la evolución temporal de las magnitudes de interés, sino que utiliza una ecuación de transición de estados, que

intenta captar la dinámica del problema relacionando las variables en un instante en función de las variables en el instante anterior. Se trata de determinar cómo evolucionan en el tiempo los parámetros que definen dicha ecuación, de forma que las diferencias entre lo observado y lo estimado se reduzcan al máximo, y la herramienta utilizada para ello es lo que se conoce como un filtro de Kalman. El Filtro de Kalman fue propuesto para sistemas dinámicos lineales a comienzos de los años 60, y está considerado como una de las herramientas fundamentales que permitieron al hombre pisar la luna, al ser empleado con éxito en el guiado de las misiones espaciales del programa Apolo.

La forma más simple de caracterizar la expansión de un virus en una población de millones de personas es mediante una progresión geométrica. De forma muy similar a la reacción en cadena de los neutrones en la fisión del átomo, si en el  $n$ ésimo día hay  $I(n)$  infectados, al día siguiente habrá un número  $I(n+1)$  dado por:

$$I(n+1) = R * I(n)$$

donde  $R$  es la **razón de la progresión geométrica**. Si  $R$  se mantuviese constante, entonces el número de infectados alcanzaría al total de la población en pocas semanas. Por ejemplo, con  $R=1.5$  (valor observado inicialmente en la propagación del virus en China), partiendo de un solo infectado llegaríamos a 38 afectados en 10 días, y a cerca de 200.000 en 30 días.

De no adoptar ninguna medida,  $R$  sólo disminuiría por agotamiento de las personas susceptibles de ser contagiadas. Afortunadamente, existen estrategias para lograr que  $R$  decrezca, siendo la más importante el aislamiento social, voluntario o forzado, de forma que  $R$  acabe alcanzando un valor inferior a uno. Para incluir en el modelo esta posibilidad, la expresión anterior se sustituye en la práctica por la de una progresión geométrica cuya razón  $R$  varía con el tiempo:

$$I(n+1) = R(n) * I(n)$$

La hipótesis de partida de nuestro trabajo es que, si bien nunca llegará a conocerse con exactitud el número de infectados reales (sólo una realización aleatoria de tests a una muestra suficiente de la población permitiría estimar ese dato), sí es posible predecir con suficiente precisión la evolución en el tiempo del parámetro  $R$ , en base a las cifras de positivos, fallecidos y recuperados reportadas por fuentes oficiales. Dicho de otra forma, resulta más sencillo estimar el cociente:

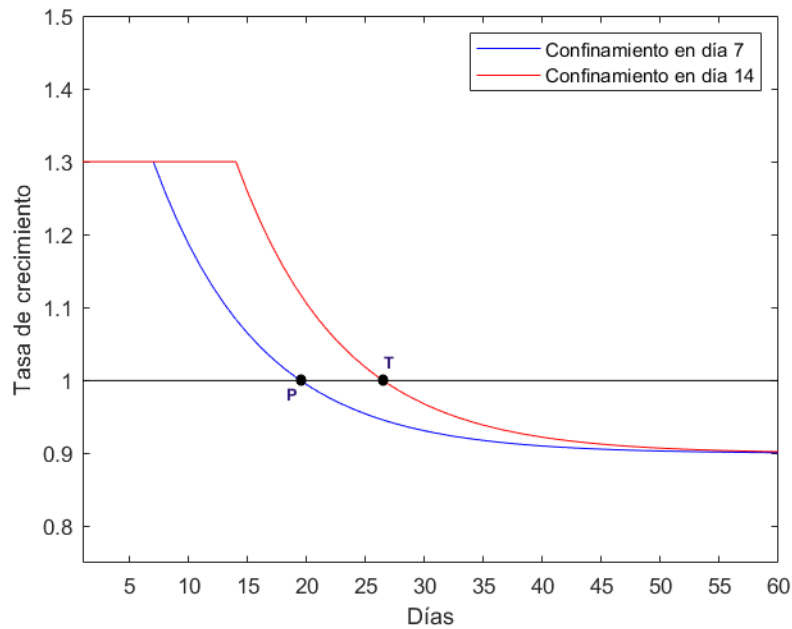
$$R(n) = I(n+1) / I(n)$$

y de ahí inferir la evolución de  $I(n)$  para un número inicial de infectados,  $I(0)$ , que estimar directamente  $I(n)$ , entre otros motivos porque  $R$  se mueve en un rango de valores bastante estrecho (menor que 2 y mayor que 0.5).

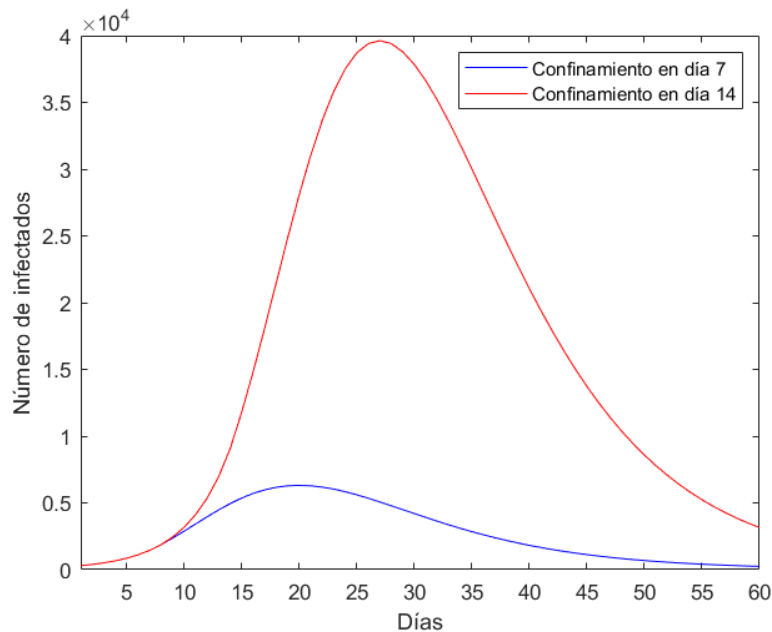
Una vez estimada la secuencia  $R(n)$ , que es el paso crítico de la metodología propuesta, puede ya deducirse automáticamente la curva de evolución de infectados, para lo cual sólo es necesario suponer un número de infectados el primer día que empieza el proceso,  $I(0)$ . Por ejemplo, si el día que se detectó el primer positivo se supone que había 4 infectados, al día siguiente habría  $I(1) = 4 * R(0)$ , el segundo día  $I(2) = 4 * R(0) * R(1)$ , el tercero  $I(3) = 4 * R(0) * R(1) * R(2)$ , y así sucesivamente.

El pico de la curva de infectados ocurrirá en un día " $p$ " tal que  $R(p)=1$ . Cuanto antes y más rápido decrezca la secuencia  $R(n)$ , antes ocurrirá el pico y menor será el número de afectados. La siguiente figura muestra un ejemplo supuesto, en el que  $R$  empieza con un valor constante 1,3 y por causas externas (un confinamiento forzado de personas), inicia un proceso paulatino de disminución hasta el valor 0,9 (menor que 1). Se consideran

dos casos, dependiendo de que el aislamiento social tenga lugar tras una o dos semanas desde el comienzo del proceso.

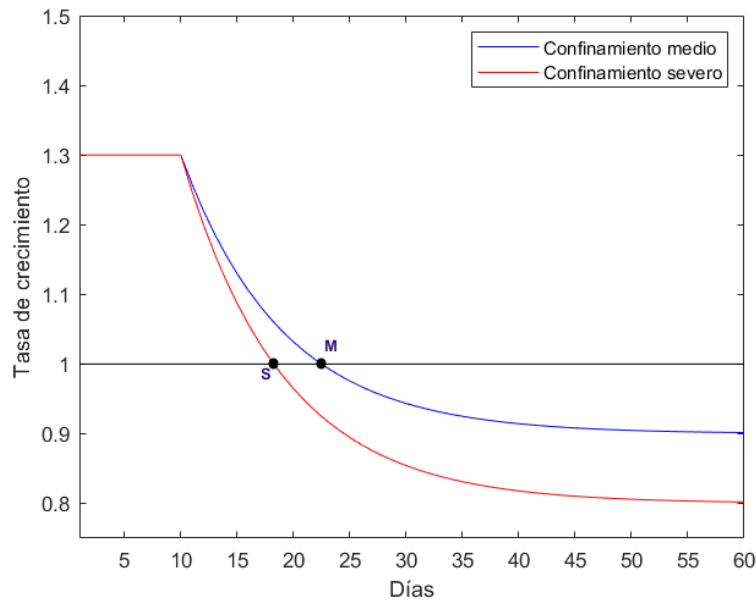


La evolución del número de infectados, suponiendo que el primer día hay 300 infectados, se muestra en la siguiente figura, donde se aprecia que, si el confinamiento empieza el día 7, el número de infectados será como máximo de unos 6.000, mientras que si el confinamiento empieza a las dos semanas, el pico ocurrirá más tarde y alcanzará los 40.000 infectados.

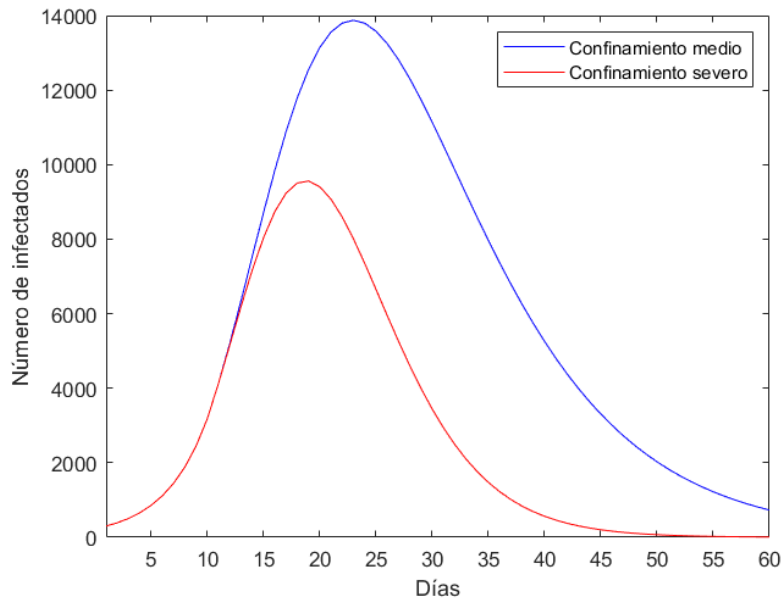


Otro factor importante, aparte del instante en que se realiza el confinamiento, es la severidad del mismo. El aislamiento social puede ser muy estricto, como el que se puso en marcha en la provincia de Hubei, en China, o puede ser más relajado, como algunos regímenes que se han visto en diversos países europeos al principio de

la epidemia. La siguiente figura muestra, para el hipotético ejemplo anterior, dos posibles evoluciones de  $R$ , hasta 0,9 (confinamiento medio) o 0,8 (confinamiento severo), en ambos casos comenzando el mismo día.



A continuación se muestra la evolución del número de afectados (suponiendo como antes 300 infectados al inicio del proceso). Con un confinamiento severo, se alcanzaría un pico de menos de 10.000 infectados, mientras que un confinamiento menos estricto conduciría a unos 14.000 personas afectadas.



En ambos ejemplos, la razón del mayor y más tardío pico se debe a que  $R$  alcanza el valor 1 más tarde. Por tanto, el objetivo es conseguir que  $R$  se reduzca lo más rápidamente posible, al menos hasta que esté por debajo de la unidad.

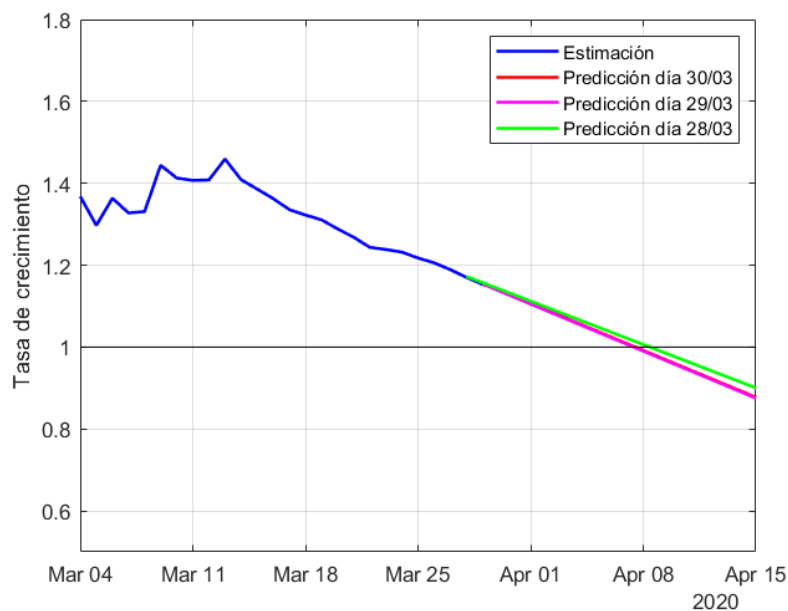
En los primeros días de la epidemia, se hizo muy popular en los medios, y hasta en círculos oficiales, una figura donde se sugería intuitivamente que el confinamiento iba a aplastar el pico, a la vez que retrasaba su aparición y lo hacía más ancho. Desde luego, el aplastamiento del pico es imprescindible para no colapsar los recursos

sanitarios, pero desde el punto de vista económico su prolongación en el tiempo es catastrófica. Afortunadamente, las gráficas mostradas anteriormente muestran que éste no es el caso. Cuanto más precoz y más severo sea el aislamiento social, antes llegará el pico, y menores serán su amplitud y duración.

Aunque están relacionadas, no debe confundirse la constante  $R$  de la progresión geométrica con la constante  $R_0$ , conocida como número reproductivo básico (número medio de personas que cada infectado contagia cuando la sociedad no toma ninguna medida), que depende de cada virus y contexto socioeconómico (para el Covid-19 se estima que  $R_0$  está comprendida entre 2 y 3).

## Resultados

La mejor estimación de la evolución de  $R$  para el caso de España, aplicando el filtro de Kalman, es la que se muestra en la siguiente figura (con los datos disponibles hasta el 30 de marzo).

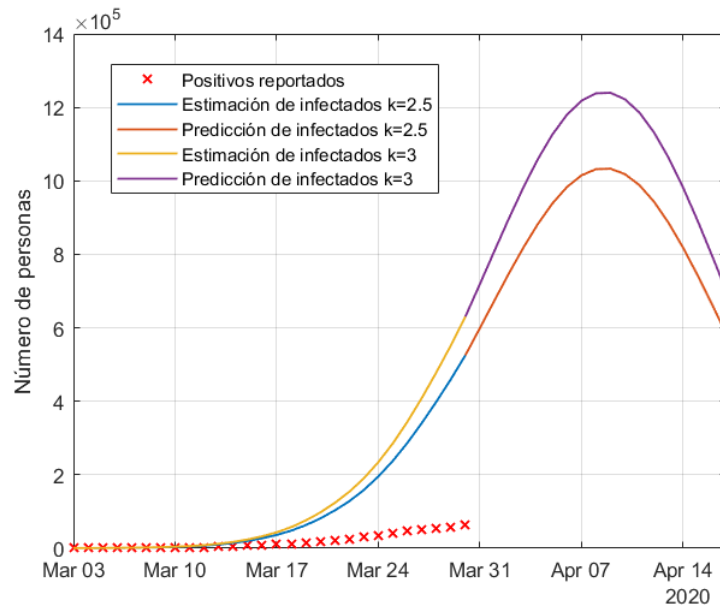


Se observa que  $R$  crecía lentamente al principio (las irregularidades que aparecen los primeros días son consecuencia de las incertidumbres del modelo y de las mayores oscilaciones de los datos de entrada). Desde aproximadamente el día 14 de marzo, fecha de comienzo del confinamiento, se aprecia claramente que  $R$  disminuye de forma sostenida. Las líneas de colores muestran una extrapolación lineal de la tendencia de  $R$ , que cruzan la línea horizontal 1 (momento del pico), alrededor del 8 de abril, aunque podría ser antes si el endurecimiento del confinamiento decretado a partir del día 30 de marzo surte efecto, como se espera.

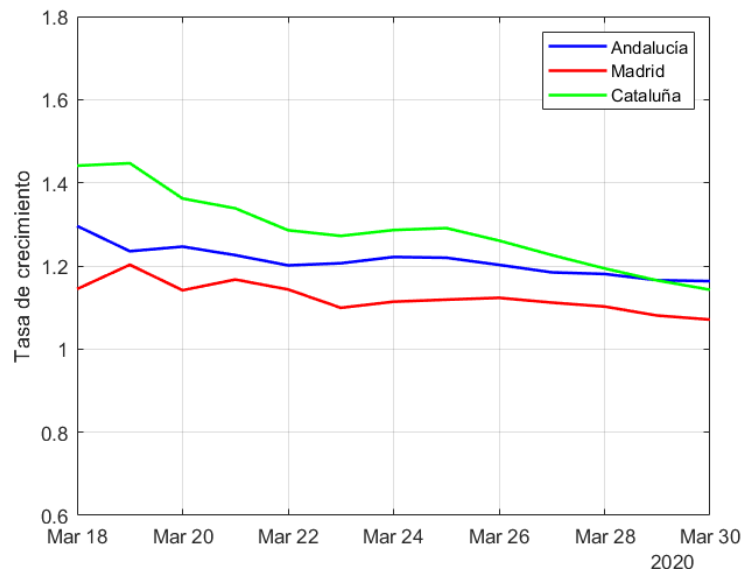
La estimación de la evolución del número de infectados, que resultaría con la  $R$  anterior, se muestra a continuación. Puede apreciarse que el número máximo de infectados estaría entre 1,1 y 1,25 millones de personas.

Es importante recalcar en este punto que, el hecho de que nos estemos acercando al pico de afectados no es incompatible con un aumento del número de fallecidos (de hecho, el pico de positivos y fallecidos diarios se producirá más tarde). Por ejemplo, 400 fallecidos en un día, sobre un total de 2.000 el día anterior, implica un 30% de incremento, mientras que, una semana después, 800 fallecidos sobre un total de 10.000 implica sólo un 8% de crecimiento. El dato de 800 fallecidos es muy importante para calibrar las necesidades de

infraestructuras, pero la disminución de R, desde 1,3 a 1,08 es lo verdaderamente relevante para averiguar a qué velocidad nos estamos acercando al pico.

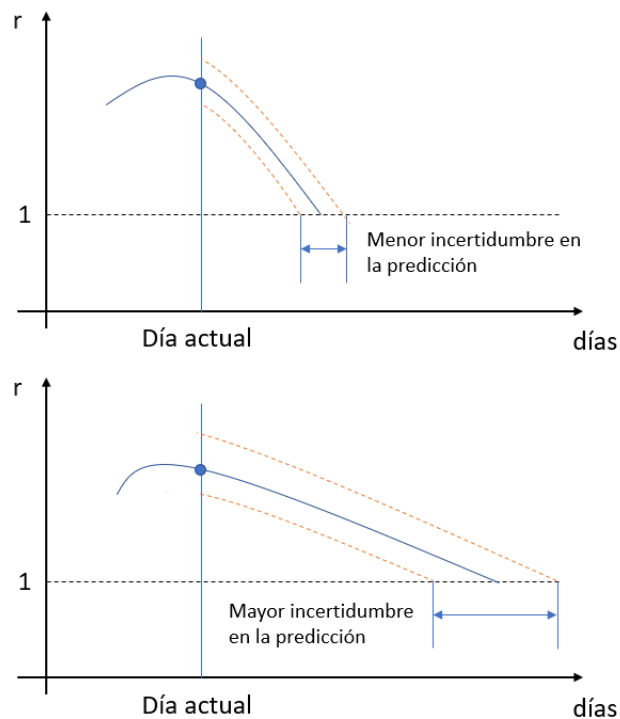


La siguiente figura muestra la evolución de R para las comunidades de Madrid, Cataluña y Andalucía. A diferencia de lo que ocurría días atrás, en estos momentos las tres comunidades autónomas parecen ir mejorando, aunque el decrecimiento de R es más acusado en Cataluña.



Finalmente, conviene insistir en que, aunque los valores estimados se muestran por conveniencia con una línea nítida y delgada, la realidad es que dichos valores llevan siempre asociados una cierta incertidumbre, cuyo impacto en la precisión con la que se estima la fecha del pico puede ser muy notable. Por ejemplo, en la parte superior de la siguiente figura se muestra la evolución de R, junto a su banda de incertidumbre (líneas rojas), en un hipotético caso en el que su valor desciende de forma rápida. Como el cruce con la línea  $R=1$  es casi

perpendicular, la incertidumbre del día en que se alcanzaría el pico es comparable a la propia incertidumbre inherente en  $R$ . En cambio, en la parte inferior, donde  $R$  decrece muy lentamente, y se aproxima a  $R=1$  tangencialmente, el momento exacto en que eso ocurre es mucho más incierto que la incertidumbre que ya de por sí arrastra  $R$ . Y no sólo eso, la fecha estimada para el pico es muy sensible a pequeñas variaciones en la tendencia de  $R$ . 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 muy elevado, de ahí la aparente discrepancia entre distintas previsiones. De hecho, si la disminución de  $R$  se llega a estancar en un valor cercano a 1, más que un pico nos encontraremos con una meseta. Esperemos que las nuevas medidas de confinamiento decretadas desde el 30 de marzo eviten esta indeseable situación.



## Próximos pasos

En estos momentos estamos aplicando la metodología desarrollada a otros países, empezando por China, donde ya se ha producido el pico, con el objetivo de confirmar la validez del estimador desarrollado y deducir patrones de evolución de  $R$ , que bien sean comunes a varios países o por el contrario sirvan para caracterizar el impacto de las diferentes modalidades de confinamiento.

El objetivo final será conseguir caracterizar la dinámica de  $R$  ante cambios en las condiciones de contorno, de forma que se pueda predecir, por ejemplo, el impacto de un levantamiento del confinamiento, dependiendo del momento en que tenga lugar y de si se hace gradual o repentinamente.