

## Proyección sobre la evolución de la incidencia del virus COVID-19

**Este informe se diferencia del homólogo en que incluye la Tabla 1b. Esta tabla muestra un tercer tipo de predicción, esta tiene en cuenta el sesgo de los valores del error observados previamente. Se proporciona también, los intervalos predictivos de confianza de este tipo de predicción.**

### Motivación

Ante la situación de alarma actual en relación al denominado COVID-19, se han puesto en marcha diferentes iniciativas que tratan de anticipar algunos de los efectos negativos que la pandemia actual genera, de esta manera es posible planificar diferentes escenarios y tomar decisiones con el objetivo de paliar las consecuencias más negativas de esta enfermedad. Así, el CEMat se encarga de coordinar las iniciativas de la comunidad matemática española relacionadas con la crisis creada por el COVID-19. Así, un grupo amplio de investigadores trata de encontrar una respuesta al problema actual de salud pública mencionado.

En particular, desde este grupo de investigación de la Facultat d'Economia de la Universitat de València, también tratamos de realizar una pequeña contribución, teniendo en cuenta las dificultades que el problema analizado presenta y que, por tanto, limita la respuesta de los diferentes modelos matemáticos que se utilizan.

### Introducción Metodológica

Existen diferentes maneras de afrontar este tipo de problemas: (i) con modelos predictivos, haciendo uso de técnicas estadísticas como los denominados GLM (Modelo Lineal Generalizado); (ii) con modelos epidemiológicos establecidos, como los denominados SIR (Susceptibles-Infectados-Recuperados) y otros modelos derivados de este; y, (iii) técnicas de Series Temporales, en las que se analiza una o más variables de interés y se establece una relación estructural de evolución temporal que se asume 'persistente' en el tiempo. (iv) Por supuesto, existen otros enfoques, pero se han descrito los más utilizados actualmente.

El grupo de investigación está trabajando en los tres tipos de análisis mencionados (i) a (iii), no obstante, este documento se centra en el tercer tipo, modelización en forma de serie temporal, ya que los resultados en el corto plazo son muy fiables y pueden ser de gran interés para realizar acciones organizativas si fuera el caso. En el Anexo 2 se describen algunos detalles utilizados en la modelización por GLM y del modelo SIR.

\* Complementariamente, los resultados de este informe se pueden consultar, junto con los análisis realizados mediante el modelo SIR y GLMs (por CCAA), a través del sitio web <https://www.uv.es/mojufan> o directamente en <https://covid-19spain.herokuapp.com/>.

## **Metodología utilizada**

En relación a la metodología utilizada, esta es ampliamente conocida en el ámbito de las series temporales. Se ha seguido el siguiente esquema:

- (1) Se consideran Tasas de Variación entre valores diarios consecutivos.
- (2) Para estimar la tendencia se inicia el proceso mediante una media geométrica móvil.
- (3) La estimación de la tendencia se realiza mediante un ajuste funcional de la serie obtenida en el paso anterior. En este caso logarítmico.
- (4) Se utiliza la estimación anterior para predecir los valores en momentos futuros.
- (5) Con los valores de tendencia y los valores observados, mediante encadenamiento, se obtienen los valores que conforman la predicción.
- (6) La predicción obtenida en el paso anterior puede ser re-estimada mediante una suavización temporal (*modificación metodológica introducida en el informe del 24 de marzo*)
- (7) Se obtiene el error de la predicción a posteriori, es decir, una vez observado el verdadero valor.
- (8) La estimación 'a priori' del error se basa en el que se introduce mediante el ajuste funcional (no aparece en este documento).
- (9) La generación de escenarios alternativos y plausibles se está diseñando y ajustando para que pueda ser utilizado en la práctica (no aparece en este documento).

### *Frecuencia de actualización del análisis*

Los resultados se amplían diariamente con los valores observados y se recalibran los parámetros del modelo y los ajustes utilizados.

### *Horizonte de predicción y variables analizadas*

Las estimaciones son útiles en el corto plazo (1-4 días), las variables analizadas son el número de fallecidos, y el número de ingresos en UCI (esta última no aparece en este documento), para el total acumulado en el conjunto del territorio nacional.

### *Datos y Fuentes de Información*

Los datos utilizados son los publicados por el Gobierno Español, aunque en la fase inicial se utilizaron datos recopilados y depurados por el grupo de trabajo *Datadista* (Github) y datos proporcionados por el *Johns Hopkins CSSE*.

## Principales Resultados obtenidos

En esta sección se presentan: (i) una tabla numérica con las diferentes predicciones realizadas en días previos, utilizando la misma metodología que en la fase de validación-no de calibración (Tabla 1) y algunos de los errores estimados (Tabla 2); y, (ii) la evolución de los valores observados de fallecimientos acumulados, junto con la predicción realizada (Figura 1); el ajuste sobre la tendencia de las tasas de variación (Figura 2); así como una comparación entre las predicciones realizadas en días previos (Figura 3).

En el anexo se pueden encontrar las estimaciones realizadas mediante suavización temporal.

## Resultados gráficos y numéricos obtenidos

En este apartado mostramos los principales resultados numéricos y gráficos para los valores estimados y para los valores estimados con suavización temporal.

La Figura 1 muestra los valores estimados a futuro con un horizonte de 4 días y obtenidos con fecha 26 de marzo.



Figura 1

Las tablas 1a-1c se derivan de la Tabla 1. Así, la Tabla 1a. muestra el último valor observado (en negro) con fecha 25 de marzo de 2020, así como las predicciones realizadas (en rojo) para los días del 26 de marzo al 29. También se incluyen las predicciones suavizadas de los mismos.

Tabla 1a.

Fecha	Obs. 25 marzo	pred 26 marzo	pred 27 marzo	pred 28 marzo	pred 29 marzo
Predicción	4089	5077	5968	6923	7931
Predicción suavizada	4089	5039	5913	6847	7931

**Como se ha comentado previamente, la Tabla 1b muestra un tercer tipo de predicción, esta tiene en cuenta el sesgo de los valores del error observados previamente. Se proporciona también, los intervalos predictivos de confianza de este tipo de predicción.**

Tabla 1b. Predicciones teniendo en cuenta el sesgo de error.

Fecha	Obs. 25 marzo	pred 26 marzo	pred 27 marzo	pred 28 marzo	pred 29 marzo
Predicción sin sesgo de error	4089	5148	6160	7324	8705
Intervalo de Predicción		[5088-5208]	[6002-6319]	[7141-7506]	[8438-8972]

Los valores desagregados de mortalidad se pueden visualizar en la Tabla 1b.

Tabla 1c. Predicciones para los valores diarios no acumulados.

Fecha	Obs. 25 marzo	pred 26 marzo	pred 27 marzo	pred 28 marzo	pred 29 marzo
Predicción	655	988	891	955	1008
Predicción suavizada	655	950	874	934	1084
Predicción corregida por sesgo de error	655	1059	1013	1163	1381

La Tabla 1 presenta los valores numéricos obtenidos para cada uno de los días en el que se ha realizado la predicción, el horizonte temporal de las predicciones varía según la cantidad de datos observados a la fecha del este informe. En este caso las predicciones mostradas no están corregidas para mantener la coherencia de la serie.

**Tabla 1. Valores observados y predicciones hasta el 29 de marzo.**

Fecha	Pred 20 marzo	pred 21 marzo	Pred 22 marzo	pred 23 marzo	pred 24 marzo	pred 25 marzo	pred 26 marzo
<b>03-mar</b>	1	1	1	1	1	1	1
<b>04-mar</b>	3	3	3	3	3	3	3
05-mar	5	5	5	5	5	5	5
<b>06-mar</b>	16	16	16	16	16	16	16
07-mar	35	35	35	35	35	35	35
<b>08-mar</b>	47	47	47	47	47	47	47
09-mar	84	84	84	84	84	84	84
<b>10-mar</b>	120	120	120	120	120	120	120
11-mar	136	136	136	136	136	136	136
<b>12-mar</b>	288	288	288	288	288	288	288
13-mar	309	309	309	309	309	309	309
<b>16-mar</b>	491	491	491	491	491	491	491
17-mar	598	598	598	598	598	598	598
<b>18-mar</b>	767	767	767	767	767	767	767
19-mar	1002	1002	1002	1002	1002	1002	1002
<b>20-mar</b>	1345	1326	1326	1326	1326	1326	1326
21-mar	1694	1700	1720	1720	1720	1720	1720
<b>22-mar</b>	2092	2106	2240	2182	2182	2182	2182
23-mar	2534	2560	2753	2834	2696	2696	2696
<b>24-mar</b>	3015	3057	3326	3457	3453	3434	3434
<b>25-mar</b>	3526	3589	3953	4151	4175	4356	4089
26-mar	4058	4147	4628	4912	4976	5191	5077
<b>27-mar</b>	4597	4720	5339	5731	5850	6102	5968
28-mar	5132	5293	6074	6597	6789	7078	6923
29-mar							7931

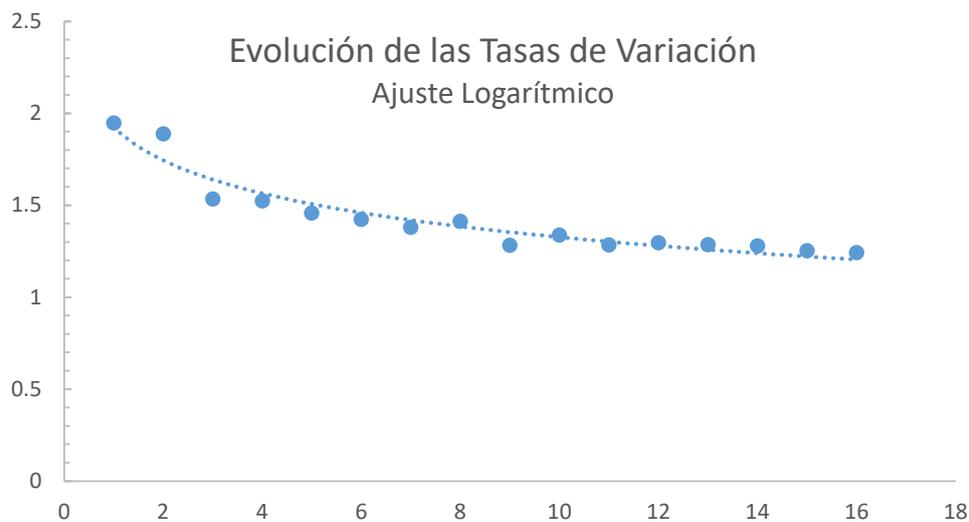
La tabla 2 muestra los errores relativos *a posteriori* que se han ido obteniendo en los días en que se ha realizado la predicción.

Tabla 2. Errores Relativos (%) por predicción y horizonte.

Error Relativo	Pred 20 marzo	pred 21 marzo	Pred 22 marzo	pred 23 marzo	pred 24 marzo	pred 25 marzo
a +1 día	1.4%	-1.2%	2.7%	5.1%	0.6%	6.5%
a 2 días	-1.5%	-3.5%	2.1%	0.7%	2.1%	
a 3 días	-4.1%	-5.1%	-3.2%	1.5%		
a 4 días	-6.0%	-11.0%	-3.3%			
a 5 días	-12.2%	-12.2%				
a 6 días	-13.8%					

La Figura 2 muestra el ajuste logarítmico utilizado para la predicción de las tasas de evolución.

Figura 2. Ajuste de las tasas de evolución.



La Figura 3 muestra el conjunto de predicciones realizadas en días previos junto con la predicción actual.

**Figura 3a. Conjunto de predicciones realizadas y datos observados (y detalle)**



**Figura 3b**



**ANEXO. Valores de predicción suavizados.**

La Tabla A1 muestra los valores de predicción para los días en que esta se ha realizado suavizada con los valores de las predicciones previas.

**Tabla A1**

Fecha	Pred 20 marzo	pred 21 marzo	Pred 22 marzo	pred 23 marzo	pred 24 marzo	pred 25 marzo	pred 26 marzo
<b>03-mar</b>	1	1	1	1	1	1	1
<b>04-mar</b>	3	3	3	3	3	3	3
05-mar	5	5	5	5	5	5	5
<b>06-mar</b>	16	16	16	16	16	16	16
07-mar	35	35	35	35	35	35	35
<b>08-mar</b>	47	47	47	47	47	47	47
09-mar	84	84	84	84	84	84	84
<b>10-mar</b>	120	120	120	120	120	120	120
11-mar	136	136	136	136	136	136	136
<b>12-mar</b>	288	288	288	288	288	288	288
13-mar	309	309	309	309	309	309	309
<b>16-mar</b>	491	491	491	491	491	491	491
17-mar	598	598	598	598	598	598	598
<b>18-mar</b>	767	767	767	767	767	767	767
19-mar	1002	1002	1002	1002	1002	1002	1002
<b>20-mar</b>	1345	1326	1326	1326	1326	1326	1326
21-mar	1694	1697	1720	1720	1720	1720	1720
<b>22-mar</b>	2092	2099	2146	2182	2182	2182	2182
23-mar	2534	2547	2616	2670	2696	2696	2696
<b>24-mar</b>	3015	3036	3132	3213	3323	3434	3434
<b>25-mar</b>	3526	3558	3690	3898	3967	4159	4089
26-mar	4058	4103	4278	4562	4666	4927	5039
<b>27-mar</b>	4597	4659	4885	5263	5410	5755	5913
28-mar	5132	5213	5500	5988	6188	6634	6847
29-mar							7931

La Tabla A2 muestra los Errores Relativos (%), hasta el 24 de marzo, por predicción y horizonte, de los valores suavizados.

**Tabla A2**

Error Relativo	Pred 20 marzo	pred 21 marzo	Pred 22 marzo	pred 23 marzo	pred 24 marzo	pred 25 marzo
a +1 dia	1.40%	-1.33%	-1.65%	-0.96%	-3.23%	1.7%
a 2 dias	-1.49%	-3.81%	-2.99%	-6.42%	-2.98%	
a 3 dias	-4.12%	-5.53%	-8.78%	-4.67%		
a 4 dias	-6.00%	-11.60%	-9.77%			
a 5 dias	-12.20%	-12.99%				
a 6 dia	-13.77%					

## ANEXO 2

Como ya se ha comentado, el análisis a corto plazo es de gran interés en el caso de ser preciso. En este documento se ha presentado una modelización mediante series temporales que está reportando datos muy precisos a corto plazo. En el transcurso de los días se verá si la ventana de predicción puede ampliarse.

En el corto plazo, y para el caso de las comunidades autónomas, se está utilizando una aproximación mediante *modelos lineales generalizados* (GLM). La elección de este tipo de modelización en el caso de CCAA es establecer diferencias de comportamiento entre las mismas, aspecto que las series temporales no permiten (al menos de una manera sencilla). Los resultados obtenidos mediante GLM son muy robustos en las predicciones (al tratar con valores esperados), y esto es útil para el caso de CCAA cuando tienen muy pocos casos observados. En estos modelos se consideran únicamente dos factores de riesgo: el número de orden del día de la observación, y que trata de capturar la evolución temporal de la variable de interés (defunciones); y la CCAA en la que se observan los casos, factor categórico, que capturará si el comportamiento es diferente según región.

Por otra parte, en el medio-largo plazo, utilizamos un modelo SIR (población susceptible, infectada y recuperada). Este modelo requiere de 4 parámetros, los que consideramos son los siguientes:

- Número de individuos Infectados inicialmente: Consideramos el valor 1.
- Número de individuos recuperados inicialmente: Consideramos que no se recupera ninguno.
- Tiempo de recuperación: utilizamos el valor que proporciona el Ministerio de Sanidad cuyo valor es de 14 días.
- Parámetro de contagio: En este caso realizamos una calibración del método mediante optimización. Para ellos se utilizan los datos observados y que proporciona el Ministerio de Sanidad, tanto en el ámbito nacional como en el regional (de CCAA).

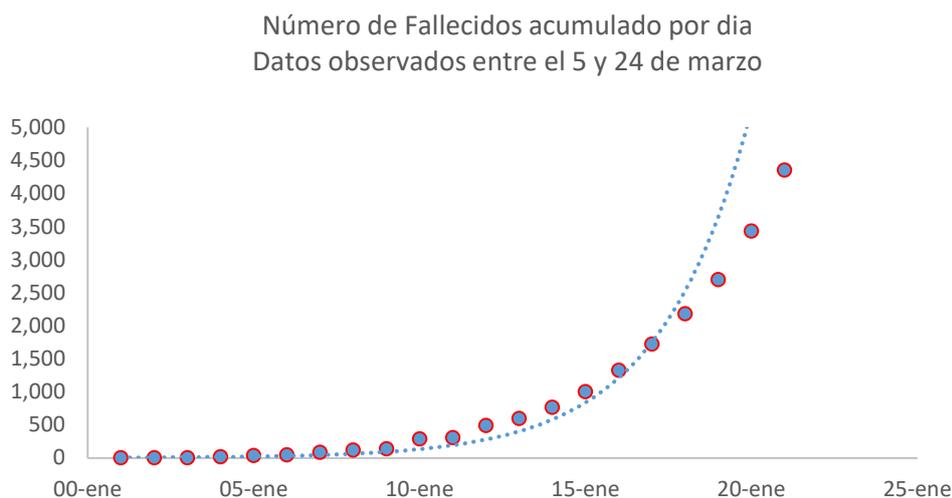
## ANEXO 3

### ¿Cómo medir la intensidad de la epidemia?

En relación al COVID-19 y, en general, a la evolución de cualquier epidemia, uno de los aspectos que más interés despierta entre los investigadores, autoridades responsables del control de la misma, y de la sociedad en general, es evaluar la intensidad de la misma diariamente. En este sentido, en esta sección se realiza un desarrollo heurístico tratando de medir la intensidad de la epidemia y su evolución temporal.

Para ello, se considera el número acumulado de personas fallecidas en el ámbito regional para el caso de España. Es habitual tratar de determinar si estos valores tienen un crecimiento de tipo exponencial o inferior. La Figura A1, muestra la serie de valores observados a fecha 25 de marzo de 2020, junto con el ajuste exponencial a los mismos.

Figura A1



Si los datos observados siguieran un modelo de crecimiento exponencial, entonces una manera sencilla que se ha valorado para medir la intensidad de la epidemia en un punto del tiempo podría reducirse en estimar la curvatura de la función en los puntos temporales considerados. Así, en dos puntos del tiempo, habría que esperar que aquel punto que presenta una menor curvatura indica que la intensidad de la epidemia es menor.

Como se observa en la Figura 1, los valores observados se desvían del modelo exponencial, por lo que la utilización de la curvatura en los puntos de la función exponencial no parece que sea inmediato.

Consideramos otras maneras de medir la intensidad de la epidemia. En particular, el uso de *Tasas de Variación Relativas* (TVR) de los datos observados; esto es:

$$r_t = \frac{X_t - X_{t-1}}{X_{t-1}}$$

La Tabla A3 presenta los datos observados entre el 10 y el 24 de marzo:

Tabla A3.

Fecha	11-mar	12-mar	13-mar	14-mar	15-mar	16-mar	17-mar	18-mar	19-mar	20-mar	21-mar	22-mar	23-mar	24-mar
TVr	0.7872	0.4286	0.1333	1.1176	0.0729	0.5890	0.2179	0.2826	0.3064	0.3234	0.2971	0.2686	0.2356	0.2737

La Figura A1 muestra estos valores. Puede comprobarse como la tasa de cambio se estabiliza rápidamente entorno a valores dentro del rango 0.21-0.31; lo que vendría a indicar que se ha reducido la incidencia de la epidemia en el transcurso de los días.

Figura A1. Serie de tasas de variación relativa

