

Treball de Fi de Grau

**Grau en Enginyeria en Tecnologies Industrials**

**Modelització i anàlisi de la COVID-19  
a partir de dades reals**

**MEMÒRIA**

**Autor:** Xavier Quintana Garriga  
**Director:** Yolanda Bolea Monte  
**Convocatòria:** Juny 2021



Escola Tècnica Superior  
d'Enginyeria Industrial de Barcelona





## Resum

Aquest Treball de Fi de Grau analitza l'evolució de la COVID-19 a Catalunya utilitzant el model de Gompertz. Mitjançant les dades reportades pel Departament de Salut s'ha ajustat la funció per analitzar l'evolució del virus des del de març fins al 17 de maig del 2020. El model ha mostrat analitzar correctament l'evolució del nombre de casos acumulats i de morts acumulades amb una  $R^2$  de 0,9984, 0,9999 respectivament.

Durant la crisi sanitària provocada per la COVID-19, no només és important conèixer les dades reals, sinó també realitzar prediccions per així les institucions sanitàries actuïn imposant les mesures corresponents. Utilitzant el programa MATLAB ha permès realitzar prediccions a curt i llarg termini, presentant errors relatius de les prediccions inferiors al 2% el primer dia, 3% el segon dia, 7% el tercer dia, 8% el quart dia i del 12% el cinquè dia. En una segona part s'ha generalitzat les prediccions per calcular l'evolució en les hospitalitzacions i les unitats de cures intensives (UCI) en la comarca del Barcelonès. En aquest cas, el model descriu la tendència amb un  $R^2$  de 0,9872 i 0,9748 respectivament.

# Sumari

<b>SUMARI</b>	<b>4</b>
<b>1. INTRODUCCIÓ</b>	<b>5</b>
<b>2. CONTEXT DEL TREBALL</b>	<b>6</b>
<b>3. MATERIALS I MÈTODES</b>	<b>8</b>
3.1. Model de Gompertz.....	8
<b>4. OBTENCIÓ DE LES DADES</b>	<b>13</b>
4.1. Algorisme d'optimització per ajustar les dades.....	13
4.2. Avaluació del model .....	13
4.3. Verificació del model .....	14
<b>5. RESULTATS</b>	<b>19</b>
5.1. Ajust del valors dels paràmetres .....	19
5.2. Prediccions a curt termini a partir del model de Gompertz .....	21
5.3. Prediccions a llarg termini a partir del model de Gompertz .....	24
5.3.1. Predicció d'altres magnituds.....	27
5.4. Estudi de la incidència .....	30
5.5. Modelització del sistema sanitari a Catalunya.....	35
5.5.1. Resultats .....	36
<b>6. DISCUSSIÓ</b>	<b>39</b>
<b>7. PLANIFICACIÓ TEMPORAL</b>	<b>40</b>
7.1. Diagrama de Gantt.....	40
<b>8. ESTUDI ECONÒMIC</b>	<b>41</b>
<b>CONCLUSIONS</b>	<b>42</b>
<b>BIBLIOGRAFIA</b>	<b>43</b>
Referències bibliogràfiques .....	43
<b>ANNEX</b>	<b>48</b>

# 1. Introducció

La ràpida propagació de la COVID-19 va provocar un repte per a les infraestructures sanitàries, col·lapsant-se en molts països. Aquest treball parteix de la necessitat de les institucions sanitàries a realitzar prediccions a curt i llarg termini per obtenir la demanda dels pacients així com la capacitat dels hospitals. Mitjançant el model de Gompertz, permet analitzar l'evolució del nombre de casos, morts, hospitalitzacions i UCI a causa de la COVID-19, i realitzar prediccions a curt i llarg termini.

En primer lloc, s'explica els mètodes emprats al llarg del projecte, el model que s'ha utilitzat, els mètodes d'estimació usats, els mètodes de validació i anàlisi del model. A continuació s'explica el programari utilitzat i el procés de sistematització de les dades per tal de fer l'estudi, validació i anàlisi de les dades de manera òptima. El treball realitzat es basa en una metodologia ja estudiada en la literatura (9) i provada en diferents casos d'estudi: Mèxic (1), Iran (2) o Espanya (3). En aquest estudi es realitza l'ajust dels paràmetres (coeficients) del model de Gompertz a partir de dades reals, i s'analitza la precisió i fiabilitat d'aquest model pel cas d'estudi de diferents parts de Catalunya. També s'estimen altres paràmetres relacionats amb el model com la incidència acumulada dels darrers 14 dies (*IA14*) o el risc de rebrot (*EPG*). Finalment, es presenten les limitacions del model i les conclusions extretes de la realització d'aquest projecte.

## 2. Context del treball

A finals del 2019 es van detectar diversos casos de pneumònia amb unes característiques desconegudes en aquell moment. Es va informar d'un brot associat a un nou coronavirus a la ciutat de Wuhan, en la província de Hubei, Xina (4). Fins a finals de gener del 2020 no s'havien publicat dades sobre una potencial transmissió de persona a persona. El 31 de desembre de 2019, els hospitals de la ciutat de Wuhan van reportar un augment del nombre de casos amb una pneumònia desconeguda (5). Seguidament, les autoritats locals van obligar tancar el Mercat a l'engròs de marisc de Huanan a causa d'una possible connexió amb el brot (6). Durant el mes de gener el nombre casos confirmats va augmentar ràpidament. El 30 de gener de 2020, l'Organització Mundial de la Salut (OMS) declarar la COVID-19 una Emergència Sanitària de Preocupació Internacional i declarada epidèmia.

El SARS-CoV-2 (síndrome respiratori agut greu coronavirus 2) és un virus de la família dels coronavirus, un tipus de virus que infecta a alguns animals i els éssers humans (5). El virus va mostrar un alt potencial de contagi entre persones, principalment es transmet d'una persona a una altra quan la persona infectada tos, parla. També és possible que en tocar una superfície amb el virus, es porti les mans a la boca, ulls o nas i es transmeti. L'alta velocitat de transmissió del virus va provocar que la comunitat internacional i els governs nacionals treballassin a generar múltiples estratègies preventives i de control. Tot i que inicialment es va aconseguir aplanar la corba epidèmica, molts països no van ser capaços de mantenir la transmissió comunitària a zero.

La possibilitat d'obtenir les dades diàries dels casos confirmats de Covid-19 ha generat una possibilitat de preveure el comportament del virus a partir de models matemàtics. Mitjançant els models matemàtics permetria adequar les mesures de control i prevenció a l'estat de la pandèmia, a més, permetria realitzar les mesures necessàries amb prou temps d'antelació.

Els models tradicionals de predicció estan dissenyats a partir d'equacions diferencials i models de predicció de sèries temporals basats en estadístiques. Els models d'equacions diferencials han d'aconseguir una equació diferencial que pugui determinar les característiques del virus a partir de les característiques del creixement demogràfic i de la transmissió entre persones (7). Mitjançant una anàlisi dinàmica del model es determina les característiques de la malaltia;

com es transmet, les tendències, les causes i factors que permeten la transmissió entre persones. A partir de les característiques obtingudes, els governs amb l'ajuda de les agències de salut pública poden dissenyar estratègies de control i prevenció. El model bàsic de predicció a partir d'equacions diferencials defineix que el nombre total de persones en una zona és constant, això permet que la transmissió sigui lògica i descriure l'evolució en el temps de diferents nodes. El mètode no és molt fiable, ja que en la realitat la població va variant i la transmissió entre les persones pot produir-se per diferents mètodes. (8) Els principals models estudiats actualment són el model SIR, el model SEIR i el model SI. Els individus del sistema es divideixen en diferents famílies, aquestes són Susceptible (S), Exposat (E), Infectat (I) i Recuperada (R). Aquests mètodes tenen com a principal problema que l'element clau en l'estudi és el número total de població susceptible, els models no estan dissenyats a canvis en restriccions dels contactes entre persones. En el cas de la COVID-19, la seva evolució varia segons les restriccions de mobilitat que imposi el govern. El virus es desenvolupa de maneres diferents segons si la regió està en un bloqueig total de la mobilitat, si les escoles estan tancades, si existeix un toc de queda, etc. A diferència dels models estudiats prèviament, el número total de persones susceptibles no és determinant, ja que el nombre de casos acumulats no s'apropen a aconseguir la immunitat del grup. Per mitigar el virus, les mesures de control han d'anar acompanyades d'un model matemàtic que permeti anticipar la dinàmica de la propagació del virus. Existeixen models fenomenològics que tenen una funció creixent, avaluen la corba de casos durant un període previ al dia  $t$  que permeti realitzar prediccions a curt termini. Entre els models fenomenològics, els models de creixement logístic són apropiats per modelar el creixement d'una pandèmia. Entre aquests models s'hi troba Gompertz (9), on el model ha sigut utilitzat per descriure satisfactòriament el comportament en altres països com Mèxic (1) o l'Iran (2).

### 3. Materials i mètodes

#### 3.1. Model de Gompertz

El model de Gompertz és conegut i utilitzat en el camp de la biologia. El principal ús és per descriure el creixement de plantes i animals, així com el nombre de bacteris i cèl·lules canceroses (10). En la literatura es troben nombroses parametritzacions del model, tot i que no s'ha realitzat un estudi de les seves propietats, el model de Gompertz-Laird és el més comú. El model de Gompertz és una variació del model de Richards de quatre paràmetres, que pertany en la família de models de creixement sigmoidals. En aquesta família s'hi troba altres models coneguts com el logístic, l'exponencial negativa o el von Bertalanffy. (11) (12) El model de Gompertz està format per tres paràmetres, on dos d'aquests paràmetres tenen efecte en la forma de la corba i un paràmetre que desplaça la corba horitzontalment sense variar la forma. Per estudiar l'evolució de la pandèmia s'utilitza un model on un únic paràmetre controla el valor inicial de la corba, el valor coincideix amb la intersecció amb l'eix y.

Una de les parametritzacions del model més comunes és:

$$W(t) = Ae^{(-e^{(-k_G(t-T_i))})} \quad (1)$$

on  $W(t)$  és el valor final en funció del temps,  $A$  defineix l'asíptota superior,  $k_G$  és un coeficient de taxa de creixement i  $T_i$  representa el temps d'inflexió. Els paràmetres que modifiquen la forma de la corba són  $A$  i  $k_G$ , en canvi el paràmetre  $T_i$

(13) A partir de l'eq.1 es troba una re-parametrització del model de Gompertz que ens servirà per realitzar l'estudi de l'evolució de la pandèmia:

$$W(t) = Ae^{(-ce^{(-k_G t)})} \quad (2)$$

en aquest cas el paràmetre  $c$  és una constant d'integració, un paràmetre d'escala del temps. Comparant els models (1) i (2) es troba la següent equació que ens permet relacionar els



paràmetres:  $T_i = \frac{\ln(c)}{k_G}$ .

Examinem un mètode per estimar la relació dels casos confirmats totals per COVID-19 amb el model de Gompertz (1825) en la població i descrivim l'evolució a curt i llarg termini de els seus paràmetres.

El mètode que utilitzem per estudiar l'evolució és el model de Gompertz. És un dels models sigmoidals més utilitzats en dades de creixement, segurament només superat pel model logístic. Inicialment el model s'utilitzava per estudiar la corba de mortalitat dels humans, però ha estat emprat en biologia per estudiar el creixement d'animals i plantes, així com el número de bacteris i cèl·lules canceroses. El model és un cas del model de Richards de quatre paràmetres i, per tant, pertany a la família de models de creixement sigmoideal de tres paràmetres de Richards, de parella amb altres models com l'exponencial negativa, el logístic o el von Bertalanffy. (14)

$$N(t) = K e^{-\ln\left(\frac{K}{N_0}\right)e^{-rt}} \quad (3)$$

El paràmetre  $K$  correspon al nombre final de casos, es pot interpretar com a l'horitzó de la pandèmia o com l'asíptota superior de la gràfica.  $N_0$  és el nombre de casos inicials en l'origen definit, finalment el paràmetre  $r$  consisteix en la ràtio de disminució del creixement exponencial inicial (15). El model s'ajusta al país o regió analitzat per tal de trobar la  $K$  i la  $r$ .

Si observem l'Eq. 3, en l'inici de la pandèmia ( $t=0$ ),  $N = K e^{-\ln\left(\frac{K}{N_0}\right)}$  la funció és una exponencial. A mesura que s'avança en el temps, el creixement s'aplana asimptòtic fins al valor  $K$  donat prèviament. La corba sigmoide és limitada en el temps, presenta un creixement monòton, i un punt d'inflexió. En aquest punt la corba passa de còncava a convexa.

Un paràmetre important a estudiar és predir quan la propagació del virus arriba a un punt crític pel que fa al nombre previst de casos totals. El punt crític que agafem és el 90% del nombre final de casos ( $K$ ), estudiar-ho permet que les institucions de govern relaxin les mesures respecte a la contenció del virus. Per calcular-ho considerem  $N(t_{90}) = 0,9K$ , sent  $t_{90}$  el dia en què es preveu arribar al 90% del nombre final de casos previst. Si calculem trobem:

$$t_{90} = - \frac{\ln\left(\frac{-\ln(0.9)}{\ln\left(\frac{K}{N_0}\right)}\right)}{r} \quad (4)$$

L'error de la predicció es calcula com:

$$\sigma_{90}^2 = \left(\frac{\partial t_{90}}{\partial K}\right)^2 \sigma_k^2 + \left(\frac{\partial t_{90}}{\partial a}\right)^2 \sigma_a^2 \quad (5)$$

on:

$$\frac{\partial t_{90}}{\partial K} = - \frac{\ln\left(\frac{-\ln(0.9)}{\ln\left(\frac{K}{N_0}\right)}\right)}{r^2} \quad (6)$$

És de major interès tenir l'habilitat de predir quan serà el nombre de casos màxims per dia. Per realitzar-ho, s'estudiarà la primera derivada de la funció de Gompertz, ja que representa els casos nous per dia:

$$\frac{dN}{dt} = rK e^{-\ln\left(\frac{K}{N_0}\right)e^{-rt}} \left(\ln\left(\frac{K}{N_0}\right)e^{-rt}\right) \quad (7)$$

Una vegada calculada la primera derivada, a partir de la segona derivada i igualant-la a zero s'obté el dia de màxim casos per dia. El punt d'inflexió permet saber quan la ràtio canvia d'augmentar a disminuir. Conèixer el punt parlant epidemiològicament és important, saber

quan la propagació està desaccelerant. (16) El punt d'inflexió és definit per:

$$t_{pic} = \frac{1}{r} \ln \left( \ln \left( \frac{K}{N_0} \right) \right) \quad (8)$$

L'error de la predicció es calcula com:

$$\sigma_{pic}^2 = \left( \frac{\delta t_{pic}}{\delta K} \right)^2 \sigma_K^2 + \left( \frac{\delta t_{pic}}{\delta a} \right)^2 \sigma_a^2 \quad (9)$$

On:

$$\sigma_{t_{pic}}^2 = \left( \frac{-\ln \left( \ln \left( \frac{K}{N_0} \right) \right)}{r^2} \right)^2 \sigma_a^2 + \left( \frac{1}{rK \ln \left( \frac{K}{N_0} \right)} \right)^2 \sigma_K^2 \quad (10)$$

Per verificar i calcular les dades del model de Gompertz es fa mitjançant una funció de regressió no lineal de Matlab. El model paramètric no lineal representa la relació entre una variable de resposta contínua i una o més variables predictores contínues de la forma

$$y = f(X, \beta) + \varepsilon \quad (11)$$

on

- $y$  és un vector  $n \times 1$  d'observacions de la variable de la resposta.
- $f$  és qualsevol funció de  $X$  i  $\beta$  que avalua cada fila de  $X$  juntament amb el vector  $\beta$  per calcular la predicció de la fila corresponent de  $y$ .

- $X$  és una matriu  $n \times p$ , amb una fila per a cada observació, i una columna per cada predictor.
- $\beta$  és un vector  $p \times 1$  de paràmetres per estimar.
- $\varepsilon$  és un vector  $n \times 1$  de pertorbacions independents i idènticament distribuïdes.

## 4. Obtenció de les dades

A l'hora de treballar en el codi de *Matlab* és necessari que les dades estiguessin en format *csv*, és per això que s'ha escollit treballar amb les dades proporcionades per la *Johns Hopkins University Coronavirus Resource Center* on s'actualitzen diàriament els fitxers de dades CSV. La informació i les dades s'han descarregat de la pàgina web de Johns Hopkins University & Medicine (17). Aquest centre format per experts en salut pública, malalties infeccioses, està enfocat en ajudar a entendre l'avanç del virus i informar el públic. Les dades són públiques i contenen taules de sèries temporals diàries. En aquestes taules s'inclou els casos confirmats, les defuncions i les recuperacions diàries, a més, estan dividides per països i regions.

### 4.1. Algorisme d'optimització per ajustar les dades

Per dur a terme l'estimació dels 3 paràmetres del model de Gompertz, és necessari aplicar un algorisme d'optimització que s'adapti a les dades reals. En el cas del model de Gompertz, es necessita un algorisme per resoldre problemes d'optimització no lineal, per així ajustar les dades en el menor temps possible. Per desenvolupar el model i realitzar tots els càlculs, es van dur a terme amb el programari d'anàlisi numèrica MATLAB. Per ajustar la regressió s'utilitza la toolbox de Matlab *Statistics and Machine Learning Toolbox*. El programa proporciona funcions i aplicacions per descriure, analitzar dades i crear models predictius de forma interactiva. Dins del programa s'utilitza la funció *fitnlm* (fit nonlinear regression model). Aquesta funció intenta trobar valors del paràmetre  $\beta$  que minimitzen les diferències mitjanes quadrades entre les respostes observades  $y$  i els predictors del model  $f(X, \beta)$ . Per realitzar-ho, necessita un valor inicial  $\beta_0$  abans de modificar iterativament el vector  $\beta$  a un vector amb un error quadrat mitjà mínim. La funció ajusta el model especificat per variables i retorna el model no lineal, per estimar els coeficients del model la funció utilitza un procediment iteratiu a partir dels valors inicials  $\beta_0$ .

### 4.2. Avaluació del model

El model utilitzat permet estimar la combinació entre paràmetres de la funció de Gompertz que millor s'ajusta a les dades reals. És més, el model permet calcular l'error entre els valors reals i els predits. Basat en els resultats obtinguts de la comparació entre els valors reals i

predits, s'utilitza un mètode quantitatiu de l'ajust del model mitjançant índex de mesura de l'error. S'utilitza la  $R^2$  per definir la correlació dels resultats i la RMSE, indicador que mesura la mida de l'error absolut (18). La qualitat del model per replicar els resultats i la proporció de variació dels resultats està definida pel coeficient de determinació  $R^2$ . La funció utilitzada pel càlcul de la precisió és:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

$y_i$  correspon al nombre de casos acumulats totals de Covid-19;  $\hat{y}_i$  és el nombre predit de casos acumulats totals de Covid-19;  $\bar{y}$  és la mitja dels casos acumulats confirmats de Covid-19. Si el coeficient de determinació s'apropa a 1, més precisa serà la predicció.

Per avaluar l'ajust del model s'utilitza altres criteris estadístics. L'error arrel-mitjana-quadrat (RMSE) és una forma estàndard de mesurar l'error en un model de predicció de dades quantitatives, es defineix com:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (13)$$

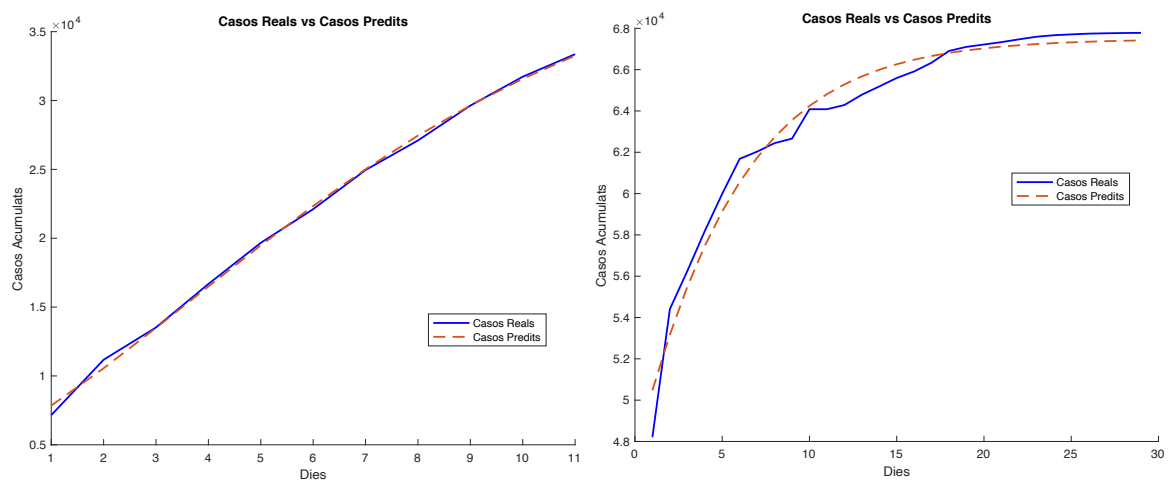
on  $n$  és el nombre d'observacions del model. El residu és una aproximació de quanta distància està el valor del model respecte als casos reals. Com més baix sigui el RMSE, el model serà més ajustat als valors reals.

### 4.3. Verificació del model

L'objectiu principal és realitzar prediccions a curt termini que siguin òptimes per així decidir sobre quines mesures s'haurien d'aplicar en els pròxims dies. Inicialment s'ha realitzat els

càlculs de previsions del model Gompertz de tres regions de la Xina a partir del nombre d'infectats acumulats totals i diaris. Les regions són Hubei, Guangdong i Henan; l'epicentre del virus va ser a Wuhan que es correspon a la regió de Hubei. (19) El desembre de 2019, els primers casos de Covid-19 van ser detectats a Wuhan, la capital de la regió de Hubei. A partir del 24 de gener, tota la província de Hubei va estar tancada, prohibint qualsevol desplaçament no essencial. El 13 de febrer, el govern local va realitzar un canvi en el protocol de diagnòstic de casos, provocant un augment exagerat del nombre de casos el 13 de febrer. Per tant, l'estudi es divideix en dues gràfiques limitant el període de l'1 al 12 de febrer, i del 13 de febrer al 12 de març.

Es realitza un estudi de les dinàmiques dels casos acumulats un cop la propagació del virus en la regió és pràcticament nul. S'utilitza aquestes regions per comprovar que el model creat es verifica i si la funció obté un bon ajust de les dades reportades. A partir de l'article (20) on s'utilitza la funció de Gompertz per predir el nombre de casos de COVID-19, es verifica el model creat amb les tres regions de la Xina estudiades. Com a exemple dels resultats obtinguts, a la Figura 1 es mostra per a les tres regions.



**Fig. 1.** Nombre total de casos a Hubei el 12 de març del 2020. En color blau i línia contínua el nombre reportat de casos, i de color vermell en discontinua les prediccions realitzades. Els resultats estan dividits en dos a causa d'un canvi en el protocol de reportar casos. S'observa els valors calculats del nombre final de casos ( $K$ ) o la ràtio de disminució del creixement exponencial inicial ( $r$ ). S'observa com la propagació del virus no segueix una distribució exponencial a partir d'un cert moment.

La Taula 1 mostra els paràmetres ajustats i els dos paràmetres que avaluen si el model s'ajusta als valors observats. Les tres regions presenten un coeficient de determinació molt proper a 1, per tant, el model de Gompertz descriu correctament l'evolució dels casos acumulats. A més s'ha verificat que els valors de la RMSE són prou menors per dir que la funció s'ajusta correctament a les dades reals. Utilitzant el model matemàtic de Gompertz s'obtenen els coeficients d'ajust  $K$ ,  $N_0$ ,  $r$ . Segons els resultats del càlcul del model, s'estima que el nombre final de casos acumulats a Hubei: 67495; Guangdong: 1355; Henan: 1279, respectivament.

L'evolució dels nous casos i l'ajust de la funció de Gompertz es mostren en la Fig. 3. En el cas dels casos reportats diàriament s'obté una funció ajustada amb una  $R^2$  elevada;  $R^2 = 0,61$  a Hubei,  $R^2 = 0,74$  a Hubei,  $R^2 = 0,97$  a Guangdong,  $R^2 = 0,97$  a Henan. Els possibles errors en l'informe dels casos a Hubei provoca que s'hagi de dividir en dues parts, és per això que la  $R^2$  no sigui tan elevada com les altres regions.

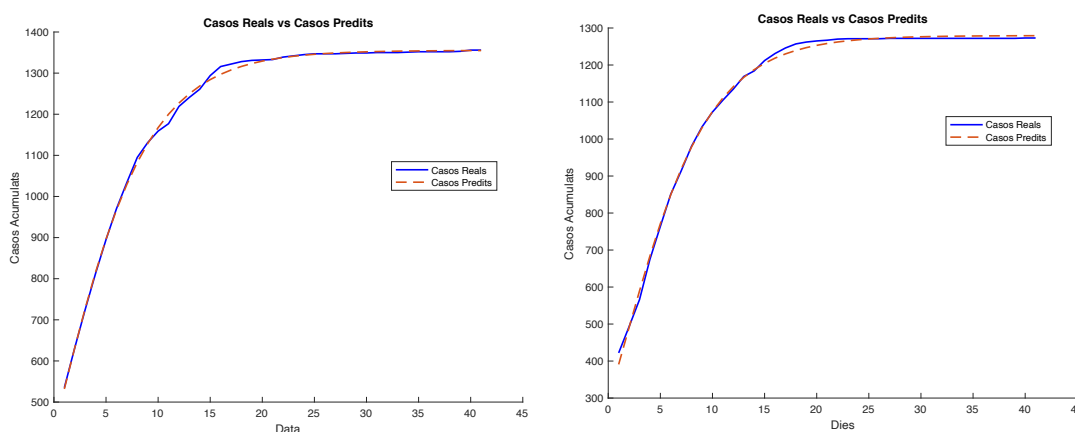
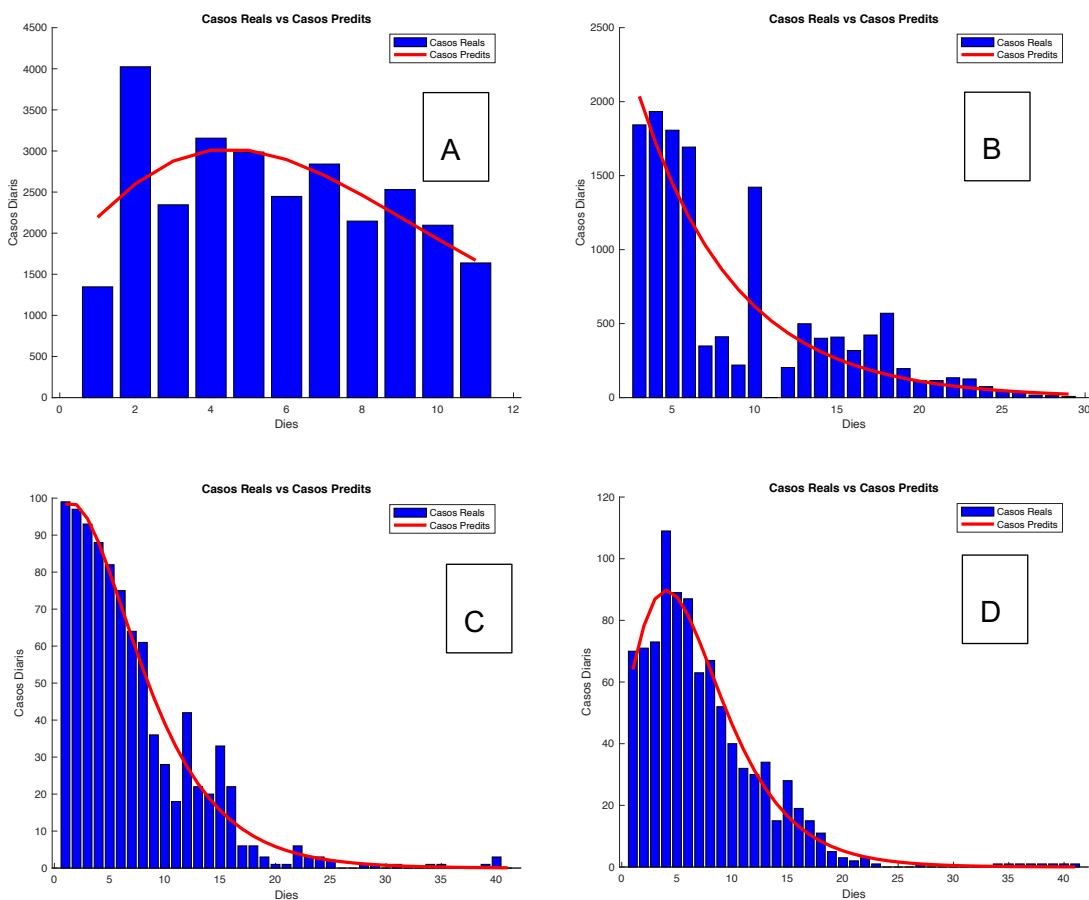


Fig.2. Nombre total de casos a Guangdong i Henan el 12 de març del 2020.





**Fig.3.** Nombre de casos diaris el 12 de març del 2020. La funció de Gompertz ajustada als casos diaris en les regions de Hubei (A) (B), Guangdong (C) i Henan (D). Les barres de color blau representen els casos diaris reportats i la línia de color vermell les prediccions realitzades.

	$R^2$	RMSE	$K$	$N_0$	$r$
<i>Hubei</i>	0,9985	378,3	42463	5473	0,1934 d <sup>-1</sup>
	0,9748	780	67495	47383	0,1970 d <sup>-1</sup>
<i>Guangdong</i>	0,9989	7,2	1355	430	0,2039 d <sup>-1</sup>
<i>Henan</i>	0,9989	9,7	127 9	295	0,2126 d <sup>-1</sup>

Taula I. Estimació dels paràmetres segons les regions.

Un cop realitzat tots els càlculs es determina que el model dissenyat és l'òptim, per tant, a partir del model matemàtic trobat podem estudiar l'evolució dels casos i morts en altres regions. Per avançar en l'estudi es treballarà en la regió de Catalunya i en la província del Barcelonès. En l'últim cas l'estudi estarà enfocat en les hospitalitzacions i en els ingressos en les unitats de cures intensives.

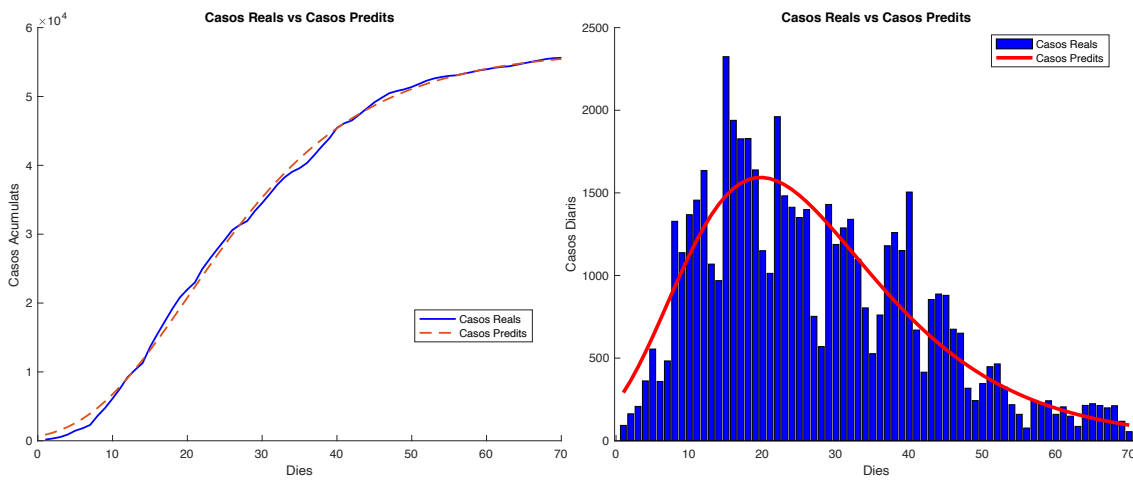
## 5. Resultats

### 5.1. Ajust del valors dels paràmetres

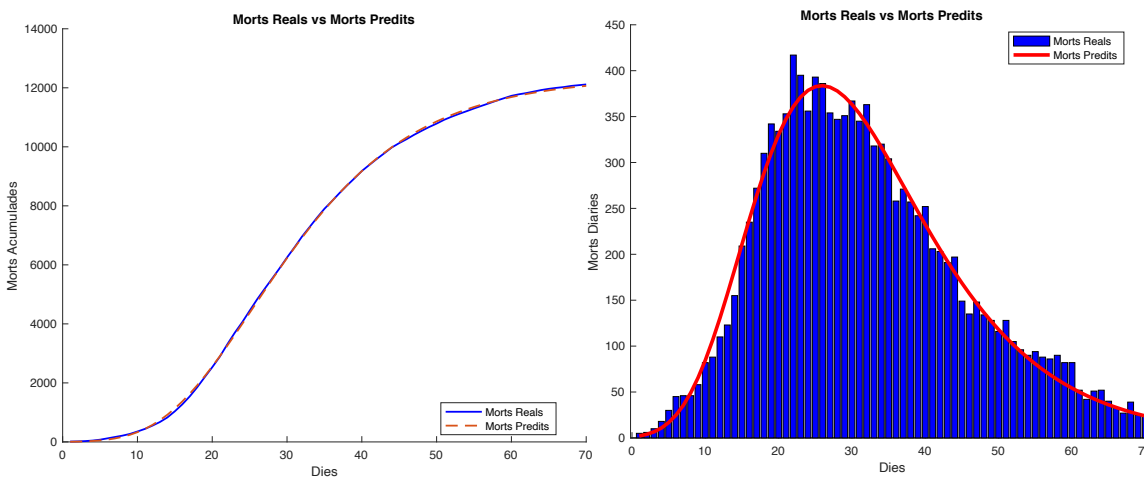
Tot i que les trajectòries de la COVID-19 comparteixen moltes propietats i característiques, cada brot té les seves pròpies, que afecten la capacitat dels governs a predir el resultat final del nombre de casos i morts acumulats. Les dades estan recollides en l'Agència de Qualitat i Avaluació Sanitàries de Catalunya (21). La Fig. 4 representa el desenvolupament dels casos acumulats i casos diaris a Catalunya, i la seva corresponent funció de Gompertz ajustada. Per analitzar el comportament s'agafa el 1r dia en què se superen els 100 casos acumulats.

En la Fig. 4 es mostra la funció ajustada en els casos confirmats totals i en els casos diaris. En la primera figura s'obté una  $R^2 = 0,9984$ , la funció s'ajusta de manera precisa als casos reportats. A més, s'obtenen els paràmetres d'ajust  $K$ ,  $N_0$  i  $r$ ; en la Taula II es mostren els intervals de confiança segons si és del 95% o 99%. En el cas del nombre de casos diaris l'ajust és relativament precís amb una  $R^2 = 0.76$ . El model està basat en la qualitat de les dades que publica les institucions, provocant errors en les dades reportades. Durant el cap de setmana es registra un nombre menor de casos i hospitalitzacions. Com a conseqüència les prediccions diàries estan influenciades per aquestes inconsistències, la diferència entre els dies laborables i el cap de setmana és significativa.

En la Fig. 5 es mostra l'evolució del nombre de morts acumulades i diàries i la relació amb els valors predits pel model de Gompertz. A més, es descriu els valors estimats pel model de Gompertz i també es presenta el coeficient de determinació. Es mostra un bon ajust ( $R^2 = 0,9999$ ) entre els valors estimats i reportats. Respecte a la precisió de les aproximacions, en el cas de la predicció dels contagis pot tenir variacions respecte als valors reportats a causa de les mesures aplicades a finals de març. En canvi, no succeeix el mateix en el model de defuncions perquè no ha tingut canvis en el seu perfil. Per tant, el model de defuncions és més robust al llarg del temps respecte al model de casos.



**Fig.4.** Ajust de la funció de Gompertz a Catalunya el 17 de maig de 2020. En el gràfic de l'esquerra està representada l'evolució del total de casos confirmats (línia blava) i la funció de Gompertz ajustada (línia discontinua vermella). En el gràfic de la dreta l'evolució dels nous casos diaris (barra blava) i la funció de Gompertz ajustada (línia vermella).



**Fig.5.** Ajust de la funció de Gompertz a Catalunya el 17 de maig de 2020. En el gràfic de l'esquerra està representada l'evolució del total de morts confirmades (línia blava) i la funció de Gompertz ajustada (línia discontinua vermella). En el gràfic de la dreta l'evolució de les noves morts diàries (barra blava) i la funció de Gompertz ajustada (línia vermella).

<b>Paràmetre</b>	<b>Casos acumulats</b>	<b>Casos diaris</b>	<b>Morts acumulades</b>	<b>Morts diàries</b>
<b><math>K</math></b>	56771	57310	12365	12262
<b><math>K</math> (CI 95%)</b>	[ 56235 – 57306 ]		[ 12329 – 12400 ]	
<b><math>K</math> (CI 99%)</b>	[ 56059 – 57482 ]		[ 12318– 12412 ]	
<b><math>r</math></b>	0,0751 d <sup>-1</sup>	0,0756 d <sup>-1</sup>	0,0832 d <sup>-1</sup>	0,0851 d <sup>-1</sup>
<b><math>r</math> (CI 95%)</b>	[ 0,0728 – 0,07741 ]		[ 0,08247 – 0,0840 ]	
<b><math>r</math> (CI 99%)</b>	[ 0,07204 – 0,07816 ]		[ 0,08222 – 0,08426 ]	
<b><math>N_0</math></b>	623	679	2,99	1,45
<b><math>R^2</math></b>	0,9984	0,7555	0,9999	0,9835
<b><math>RMSE</math></b>	769,2	293,3	47,65	16,6

Taula II. Estimació dels paràmetres el 17 de maig de 2020

## 5.2. Prediccions a curt termini a partir del model de Gompertz

Un dels principals objectius de realitzar una funció matemàtica és realitzar prediccions a curt i llarg termini per així els governs tinguin una previsió del futur. Mitjançant aquestes prediccions tenen la capacitat de modificar les mesures de control en la contenció del virus. Amb les dades obtingudes a Catalunya es realitza un estudi del comportament de la funció de Gompertz a través del temps.

Un cop ajustada la funció i calculat els paràmetres  $K$ ,  $N_0$  i  $r$ , es poden realitzar prediccions a curt termini. En les Taules III/IV es mostren les prediccions realitzades pels cinc dies següents, així com el seu interval d'error de les prediccions.

<i>Dia</i>	<i>Nº de casos totals predits</i>	<i>Interval</i>	<i>Nº de casos totals reals</i>
18/5/2020	55546	[ 55140 - 55952 ]	55843
19/5/2020	55634	[ 55221 – 56047 ]	56091
20/5/2020	55715	[ 55296 – 56135 ]	56307
21/5/2020	55791	[ 55366 – 56216 ]	56581
22/5/2020	55861	[ 55430 – 56292 ]	56847

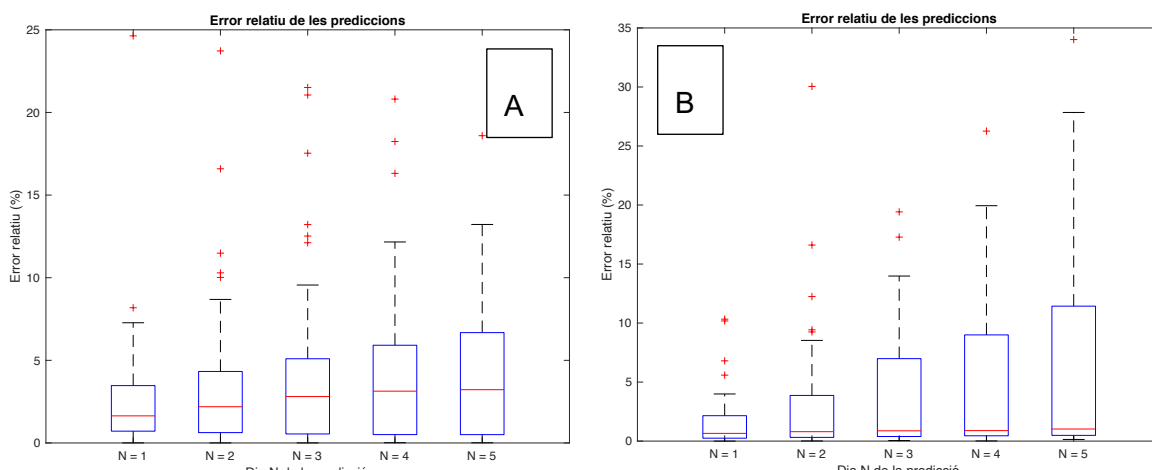
*Taula III. Prediccions dels casos acumulats a Catalunya dels següents 5 dies.*

<i>Dia</i>	<i>Nº de morts totals predits</i>	<i>Interval</i>	<i>Nº de morts totals reals</i>
18/5/2020	12088	[ 12062 - 12115 ]	12130
19/5/2020	12110	[ 12083 – 12137 ]	12160
20/5/2020	12130	[ 12103 – 12158 ]	12191
21/5/2020	12149	[ 12121 – 12177 ]	12217
22/5/2020	12166	[ 12137 – 12195 ]	12239

*Taula IV. Prediccions de les morts acumulades a Catalunya dels següents 5 dies.*

S'utilitza un interval de confiança del 99% i s'estudia l'error relatiu de la predicció respecte al nombre exacte, i si la quantitat reportada està dins de l'interval de l'error de la predicció. Si el valor reportat per les fonts oficials cau dins de l'interval previst, considerem la predicció correcta, en el cas contrari, es considera erroni. L'error relatiu es determina com  $\text{abs}((\text{valor reportat} - \text{predit})/(\text{valor reportat}))$ . Per realitzar les prediccions s'agafa les dades reals fins un dia determinat i executem el programa de Matlab per obtenir els resultats de les prediccions

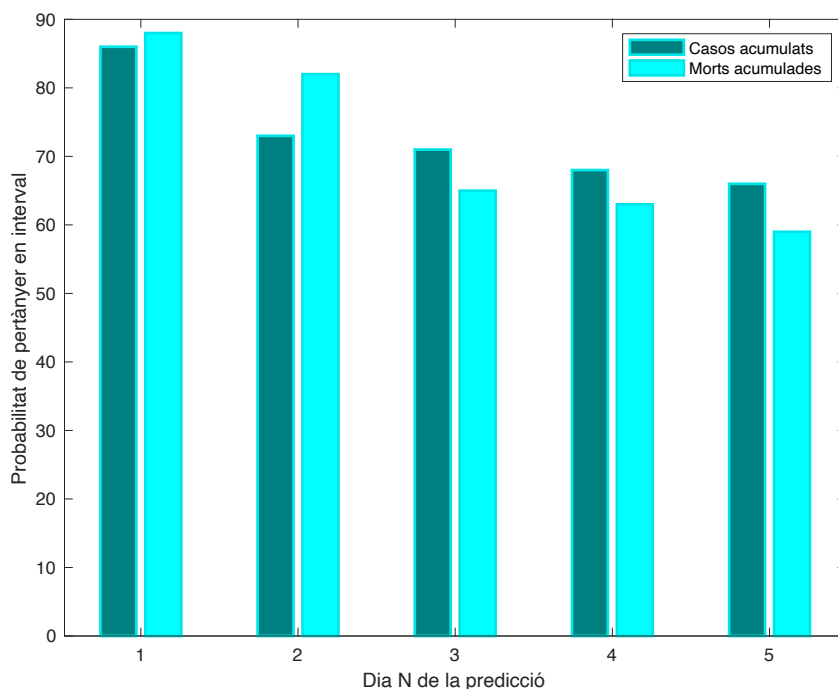
pels 5 dies següents. Els dos paràmetres estudiats permeten determinar la fiabilitat de les prediccions, i calibrar les barres d'error del model i calcular el percentatge d'èxit. S'utilitza un gràfic de barres per definir la probabilitat de pertànyer en l'interval, i un boxplot per definir l'error relatiu. (22) El boxplot permet representar les dades a través dels quartils, que són conjunts que quantifiquen la concentració dels valors, i que ordena les variables de menor a major i les divideix en grups amb el mateix nombre d'observacions.



**Fig.6.** Error relatiu de les prediccions. Anàlisi de la precisió per a prediccions de fins a 5 dies en els casos acumulats (A) i en les morts acumulades (B).

En la Figura 6 trobem l'error relatiu de les prediccions respecte als valors reals. S'obté un error inferior al 2% per les prediccions realitzades pel següent dia. Per a prediccions de 2 i 3 dies s'obté un error inferior al 3%, finalment per a 4 i 5 dies prediccions s'obtenen errors entre un 3 i 4%. En el cas de les morts acumulades s'obtenen errors relatius d'1% pràcticament els cinc dies. S'observa que la caixa és menys simètrica a mesura que s'augmenten els dies, existeix una asimetria positiva o segada a la dreta. Les dades són més disperses i el rang augmenta, tot i que la mitjana es manté pràcticament constant.

A continuació s'avalua la probabilitat de pertànyer en l'interval de les prediccions realitzades cada dia. Es va realitzar un estudi d'ençà que es van sobrepassar els 100 casos acumulats fins que es van reduir els casos diaris a menys de 1000. S'obté al voltant d'un 85% de confiança per les prediccions per l'endemà, la probabilitat descendeix en els següents dies fins a obtenir un 60% de probabilitat. En el cas de les morts, la probabilitat en el primer dia és pràcticament el 90%. En el segon dia la probabilitat disminueix tot i que es manté per sobre del 80%, en els següents dies la probabilitat disminueix per sota del 70%.



**Fig.7.** Probabilitat que les prediccions realitzades cauen en l'interval de predicció per cada dia.

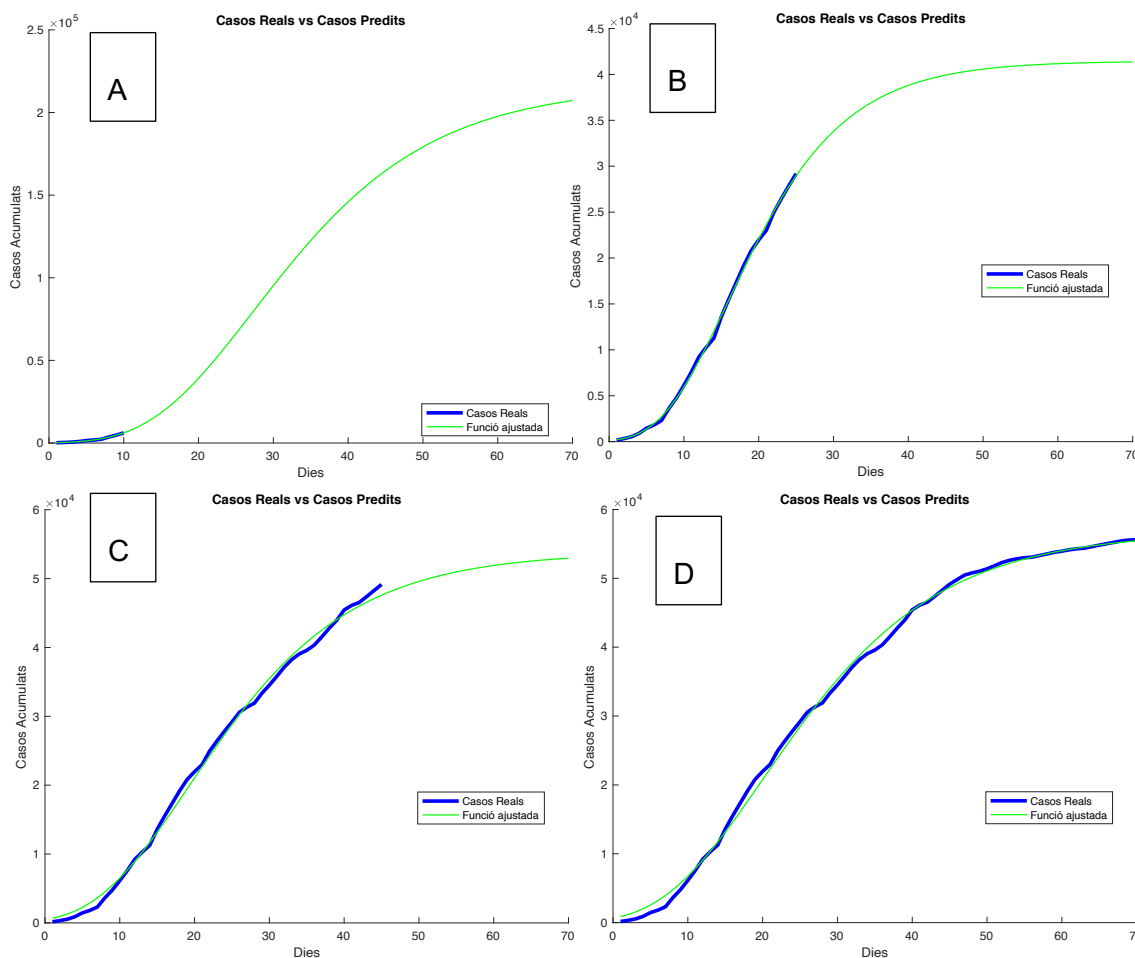
Obtenir els errors relatius respecte a les prediccions dels següents dies és necessari per entendre el comportament de l'epidèmia, però també és necessari definir predir l'evolució durant. Si tenim una predicció correcta del comportament des de l'inici, permetrà que els governs actuïn prèviament amb les mesures de control i prevenció correctes. Ajustant el model de Gompertz diàriament permet calcular diferents paràmetres que determinen l'evolució. Aquests paràmetres són la ràtio de disminució del creixement exponencial inicial i el nombre final de casos acumulats.

### 5.3. Prediccions a llarg termini a partir del model de Gompertz

En aquest apartat s'elabora una anàlisi de la precisió de les prediccions realitzades prèviament a Catalunya. Es realitza un estudi dels casos acumulats totals i de les morts acumulades. A l'hora de calcular-ho, s'emmagatzema les prediccions realitzades cada dia i les seves corresponents variables. Executem el programa des del 9 de març fins al 17 de maig de 2020,



i comparem les prediccions amb el nombre de casos reals reportats per les fonts oficials. La forma bifàsica de la corba és a causa que en l'inici de l'epidèmia la majoria de casos són importats d'altres països, és per això que el nombre no augmenta. A mesura que passen els dies la transmissió local s'incrementa provocant la que la corba sigui exponencial. Les mesures de prevenció i control fan que la corba sigui pràcticament asimptòtica.

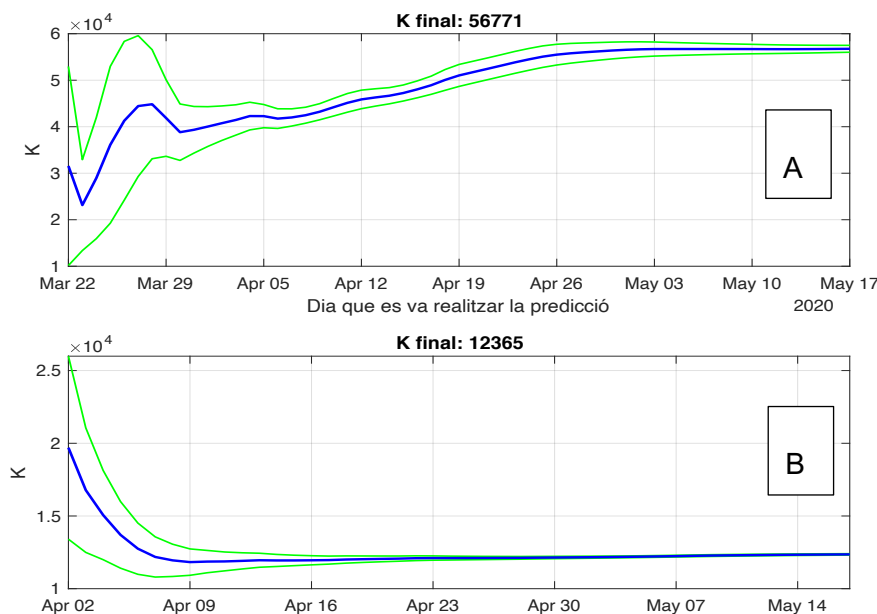


**Fig.8.** Ajust de la funció de Gompertz a Catalunya el 10 de març (A), el 25 de març (B), el 14 d'abril (C) i el 17 de maig de 2020 (D). La línia blava representa el nombre de casos acumulats i la línia verda la funció ajustada.

En la Fig. 8 s'analitza l'evolució de la funció de Gompertz ajustada en funció del temps que s'analitza. S'ajusta la funció així com els seus paràmetres en quatre moments segons el nombre de casos acumulats reportats prèviament. D'aquesta manera, permet calcular

l'evolució dels casos en els pròxims dies. Els valors dels tres paràmetres varien respecte als valors finals calculats en la Fig. 4; a mesura que avança, el nombre final de casos  $K$  disminueix. En les dates estimades pel fi de la pandèmia, s'ha de tenir en compte que són aproximacions i que si s'apliquen mesures de control extremes, poden tenir alteracions en el comportament del model. Per exemple, el confinament de la població va provocar que el nombre de casos i morts disminueixi considerablement a partir de la primera setmana.

En les Fig. 9/10 es mostra l'evolució dels paràmetres  $K$  i  $r$  en els casos i morts acumulats. En l'inici l'interval de confiança és ampli, però a mesura que s'avança el valor convergeix al seu valor final i l'error en l'interval de confiança disminueix. Per exemple, a principis d'abril el model predeia quasi 20.000 morts a Catalunya, i a mesura que passaven els dies el valor va disminuir fins a situar-se sobre les 12.000 morts. A partir de principis d'abril, l'estimació es manté constant al llarg del temps. Aquesta estabilitat dels paràmetres serveix com a mesura de robustesa en l'estimació de morts. En el cas dels contagis, l'evolució del seu horitzó varia bastant entre l'inici i el final. Aquestes variacions demostren que les prediccions a llarg termini són difícils de realitzar.



**Fig.9.** Evolució de les prediccions a llarg termini. Evolució de la predicció del nombre final de casos confirmats totals (A) i evolució del nombre final de morts totals (B). Càlcul dinàmic del paràmetre  $K$  en blau fosc, de color verd les

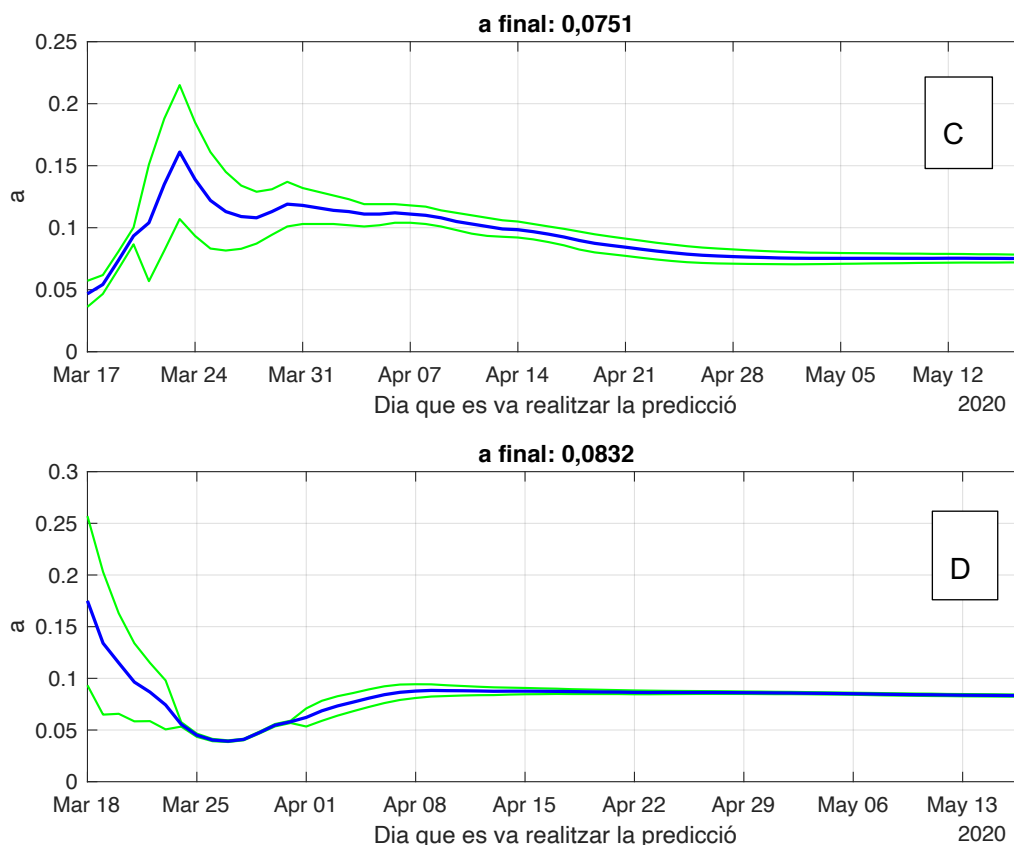


Fig.10. Evolució de la  $r$  a Catalunya fins al 17 de maig de 2020. En el nombre final de casos totals (C) i en el nombre final de morts totals (D).

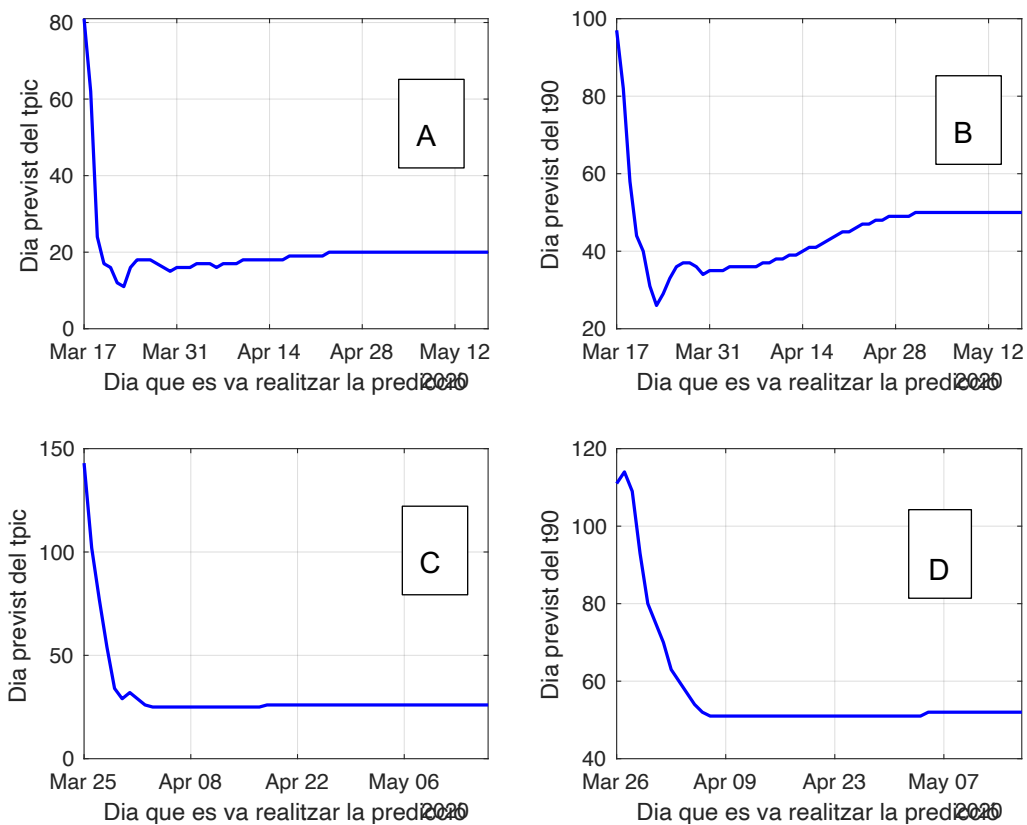
### 5.3.1. Predicció d'altres magnituds

Per aconseguir una predicció més consistent i sòlida, s'utilitza l'ajust del model en el nombre de casos acumulats i diaris. Es pot estudiar quan s'obindrà el pic de casos diaris, i el temps necessari per arribar al 90% de casos acumulats. Estudiant aquests dos paràmetres, permet realitzar prediccions a llarg termini i estudiar diàriament la dinàmica de l'epidèmia. S'estudia l'evolució del  $t_{pic}$  i  $t_{90}$  entre el 17 de març i el 17 de maig de 2020.

$$t_{90} = - \frac{\ln\left(\frac{-\ln(0.9)}{\ln\left(\frac{K}{N_0}\right)}\right)}{r} \quad (14)$$

El punt d'inflexió permet saber quan la ràtio canvia d'augmentar a disminuir. Conèixer el punt parlant epidemiològicament és important, saber quan la propagació està desacelerant. Per estimar el  $t_{pic}$  s'utilitza l'Eq. 8 definida prèviament.

Es calcula diàriament els paràmetres ajustats de la funció de Gompertz i l'evolució dels paràmetres explicats prèviament, en la Fig. 11 es defineix el  $t_p$ ,  $t_{90}$ . S'observa com les prediccions inicien amb un error significatiu, però a mesura que s'avança els valors convergeixen al valor calculat l'últim dia. En l'inici de l'estudi l'epidèmia estava en la seva fase inicial, és per això que hi ha un error elevat en els primers dies. En el cas de les morts, l'error en el valor inicial és significatiu, però a mesura que avancen els dies el valor coincideix amb el valor final. L'error que s'obté en els primers dies és degut al fet que l'estudi està realitzat en l'inici de la pandèmia, durant aquest període creix de forma exponencial provocant errors grans.



**Fig. 11.** (A) i (C) Convergència de la predicció del  $t_{pic}$  del nombre màxim de casos i morts diaris respectivament; (B) i (D) convergència de la predicció del  $t_{90}$  del nombre total de casos i morts respectivament. L'inici es correspon al 9 de març de 2020.

	Casos		Morts	
	Predicció	Reportat	Predicció	Reportat
$t_{pic}$	29/3/2020	23/3/2020	3/4/2020	30/3/2020
$t_{90}$	28/4/2020	25/4/2020	30/4/2020	29/4/2020

Taula V. Comparació de la situació real amb la predicció.

En la Taula V es mostra la data prevista que el nombre de casos diaris confirmats arriba al màxim i el 90% dels casos acumulats. Segons la predicció del model s'arriba a finals de març de 2020 i finals d'abril 2020 respectivament. Els resultats són consistents tot i que existeix un error. Si s'aplica el model en les morts, s'obté un error inferior respecte els casos diagnosticats.

## 5.4. Estudi de la incidència

Una vegada calculada el dia del nombre màxim de casos, i el  $t_{90}$ , com s'ha definit prèviament, es pot comparar els dos paràmetres temporals amb la incidència acumulada, que és, el nombre de casos per 100000 habitants.

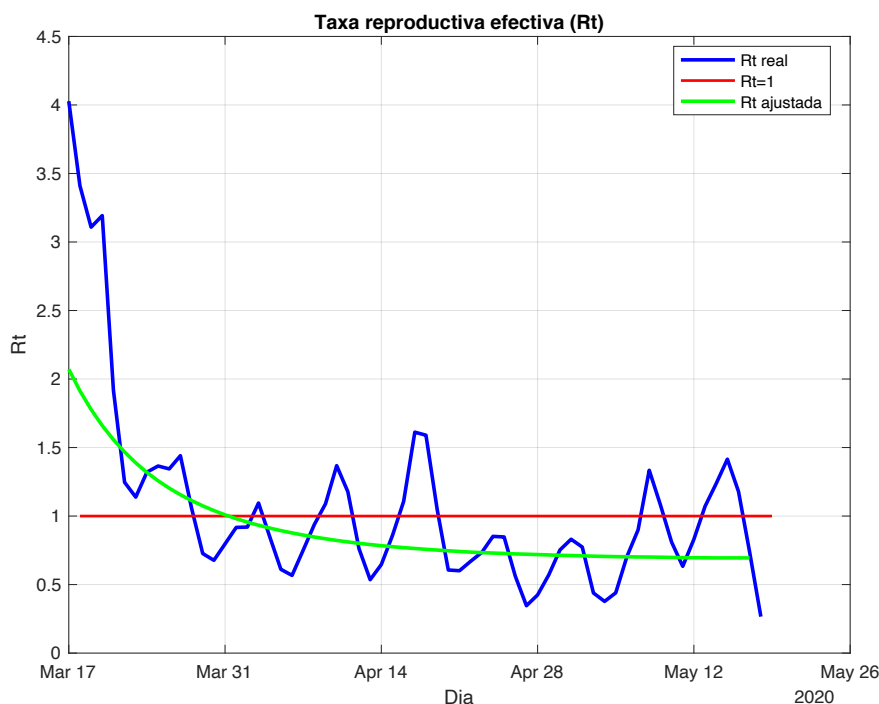
La  $Rt$  és una mesura empírica per avaluar la velocitat de propagació de la pandèmia. La mesura determina el nombre mitjà de persones infectades per una persona infecciosa. Pel dia anterior s'avalua de la següent manera:

$$Rt(t-1) = \frac{N(t) + N(t-1) + N(t-2)}{N(t-5) + N(t-6) + N(t-7)} \quad (15)$$

On  $N(t)$  són el nombre de casos nous confirmats el dia  $t$ .

És necessari remarcar que la precisió de la predicció es basa en el número de reproducció bàsic per als models estudiats prèviament. El número bàsic de reproducció és fonamental per determinar la transmissibilitat entre la població. En l'inici de l'epidèmia, la  $Rt$  és dinàmica, depèn de les mesures de prevenció i control que apliquin els governs nacionals i locals. A causa de la variabilitat de la  $Rt$  en les etapes de l'epidèmia, és difícil realitzar una predicció correcta sobre la transmissió del virus mitjançant models matemàtics que divideixen la població en compartiments; per exemple,  $SIR$ ,  $SEIR$ , etc. S'utilitza el model de Gompertz per predir el comportament, ja que el model permet utilitzar les dades històriques dels casos, i

alhora utilitzar les dades en temps real.



**Fig.12.** Velocitat de reproducció a Catalunya el 17 de maig de 2020. La línia blava representa la  $R_t$  real, la línia vermella la predicció realitzada. La línia verda marca  $R_t = 1$ , que indica que la propagació s'alenteix.

La mesura presenta variacions clares en el dia a dia a causa del retard i de les irregularitats en els informes presentats. A més, s'ha d'afegir que en els caps de setmana a causa de la jornada laboral el nombre de casos reportats disminuïa respecte als dies laborables. Això provocava que els endarreriments dels casos es reportessin en els dies següents. És per això que és necessari realitzar una mitja de les dades de set dies per així obtenir un quadre complet de la situació de la regió. (23) Definim la velocitat de reproducció dels últims set dies com  $\rho_7$ , i es defineix la mitjana:

$$\rho(7) = \frac{1}{7} \sum_{i=-3}^3 \rho_{t+i} \tag{16}$$

El segon component és la incidència acumulada dels darrers 14 dies (IA14), permet estimar quantes persones hi ha amb capacitat de contagiar la malaltia en un moment determinat. Es calcula com la suma del nombre de nous casos del dia actual més els casos dels 13 dies anteriors per 100000 habitants.

Després de calcular-ho ens permet definir el risc de rebrot o EPG (índex de creixement potencial). Consisteix en el producte dels dos components anteriors, la velocitat de reproducció mitjana dels darrers 7 dies ( $\rho7$ ) i la incidència acumulada dels darrers 14 dies (IA14).

$$EPG = \rho7 * IA14 \quad (17)$$

L'indicador permet valorar el creixement de la pandèmia, mesura els possibles nous casos diagnosticats els 14 dies vinents i està relacionat amb la probabilitat d'aparició de nous brots epidèmics. Per aconseguir una EPG baixa, es necessita un valor baix de velocitat i nombre baix de casos actius. Si la IA14 és gran i la  $\rho7$  és baixa o si la  $\rho7$  és gran amb una IA14 baixa, són situacions potencialment amb risc. Quan els dos components IA14 i  $\rho7$  són grans, representa que hi havia una població nombrosa amb el virus i que es va propagar prèviament. En la Fig. 13 es mostra l'evolució de la IA14 mitjançant l'ajust de la funció de Gompertz. L'error relatiu obtingut és de 9,13%.



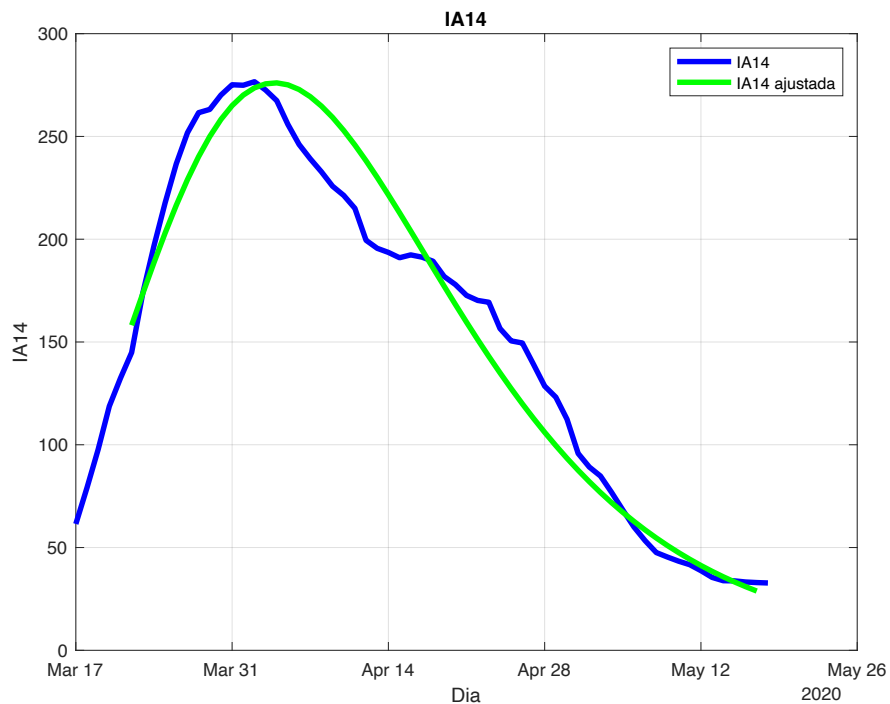


Fig. 13. Evolució de la incidència acumulada dels darrers 14 dies (IA14) a Catalunya del 17 de març el 17 de maig de 2020.

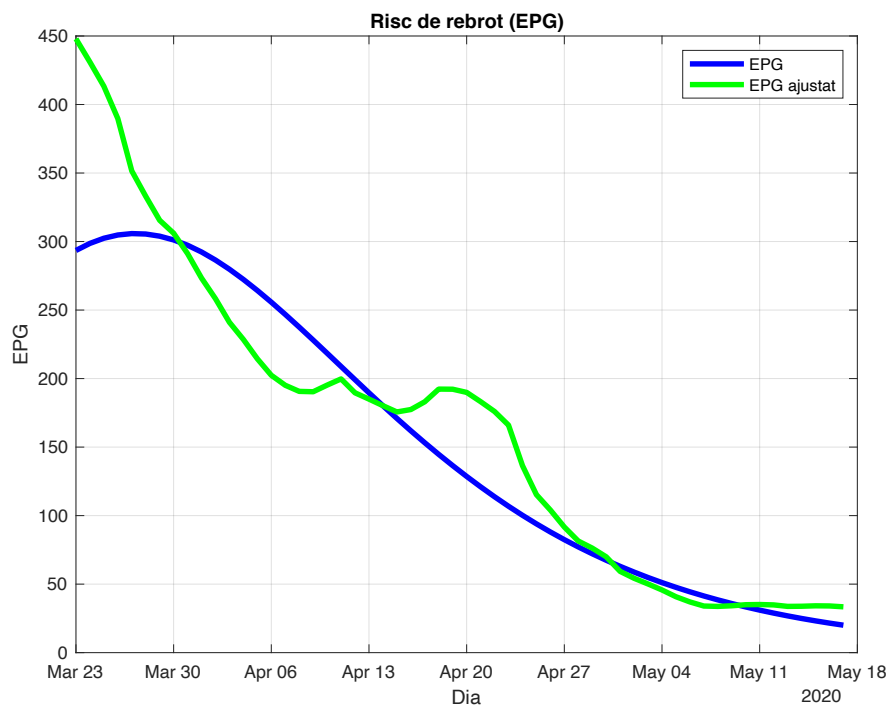


Fig. 14. Evolució del risc de rebrot a Catalunya fins el 17 de maig de 2020.

Després d'ajustar la funció de Gompertz a partir dels casos acumulats, en les Fig. 13/14 es prediu la incidència acumulada dels darrers 14 dies (*IA14*) i l'evolució del risc de rebrot a Catalunya.

Una de les maneres per veure la dinàmica de les gràfiques és comparant-les amb la funció ajustada al nombre de casos diàries. El comportament dels paràmetres és similar, en el cas de la *IA14* que necessita els casos dels últims 7 dies té un retard respecte al nombre de casos diaris. S'observa com el seu pendent augmenta ràpidament i en pocs dies, quan decreix és el contrari doncs, necessita més dies per tornar als valors de l'inici de l'epidèmia. Ajustant la funció als casos acumulats permet definir el comportament de la incidència acumulada de forma precisa.

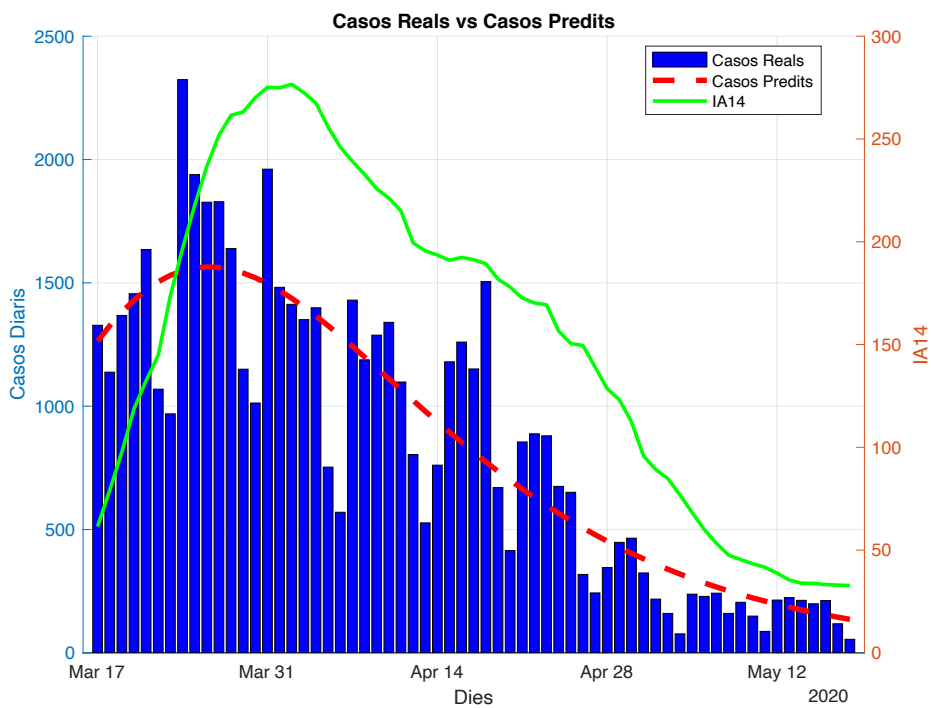


Fig.15. Comparació entre la *IA14* i el nombre de casos diaris reportats.

## 5.5. Modelització del sistema sanitari a Catalunya

La COVID-19 ha suposat un repte pel sistema sanitari, ja que han hagut d'adaptar-se i investigar en noves tècniques per no tenir saturació en la demanda de llits hospitalaris i en la unitat de cures intensives. El període mitjà d'incubació de la COVID-19 és de 5 dies i el 97,5% dels que presenten símptomes els tindran abans dels 12 dies de l'inici del contagi (24). Els principals símptomes són febre, dificultat respiratòria (dispnea), tos seca i símptomes gastrointestinals, tot i que és possible que no presentis cap signe de febre durant el contagi. En el cas de les hospitalitzacions el període mitjà des de l'aparició dels primers símptomes és de 7 dies i de la insuficiència respiratòria de 8 dies. La mitjana de la durada a l'hospital fins a la recuperació és de 10 dies. Dins de les hospitalitzacions, el 25% dels pacients entra en la unitat de cures intensives. (25) La taxa de mortalitat oscil·la entre el 2 i 3%, però en el cas dels adults amb patologies prèvies pot arribar a un 11%.

A causa del brot del virus, ha provocat que el sistema sanitari estigui col·lapsat. Per reduir l'estrès s'ha realitzat millores en l'eficiència de la distribució dels pacients. Aquesta situació ha provocat mesures extraordinàries, ja que els recursos de les organitzacions sanitàries són limitats. A l'hora de definir aquestes mesures, la capacitat de tenir una funció matemàtica és una eina notable per realitzar prediccions en la capacitat hospitalària i en la unitat de cures intensives.

Després de realitzar un estudi utilitzant la funció de Gompertz per predir el nombre de casos per COVID-19 i el nombre de morts, es realitza la mateixa funció per estudiar l'evolució en el nombre d'hospitalitzacions i en la unitat de cures intensives (UCI). A més, permetrà realitzar prediccions a curt termini necessàries per a les organitzacions sanitàries.

S'utilitza el mateix codi de MATLAB implementat en els apartats anteriors. Per implementar el model, s'adquireix les dades a partir de les fonts oficials de Catalunya (21), que són actualitzades cada dia. S'utilitza la mitjana de 7 dies, tres dies abans, tres dies després i el dia calculat, per així evitar els possibles errors en la publicació de les dades del cap de setmana. A l'hora de seleccionar les dades només es tenen comptabilitzades les dades a partir de l'1 de maig de 2020. Sense tenir les dades prèvies a aquesta data, no es pot realitzar un estudi complet de l'evolució de la COVID-19 en la 1a onada. S'ha d'afegir que les dades reportades no són acumulades, per tant, l'estudi haurà de ser a partir de les dades diàries.

En la 1a onada s'utilitzen les dades del nombre de pacients ingressats als hospitals per COVID-19 i el nombre de pacients ingressats a l'UCI amb COVID-19 en l'última data del període. El nombre de pacients ingressats a l'UCI ja estan inclosos en els ingressats. El model serà ajustat per realitzar prediccions dels valors diaris fins que la variació del nombre d'hospitalitzacions sigui menor de 6, i de 2 en la UCI.

### 5.5.1. Resultats

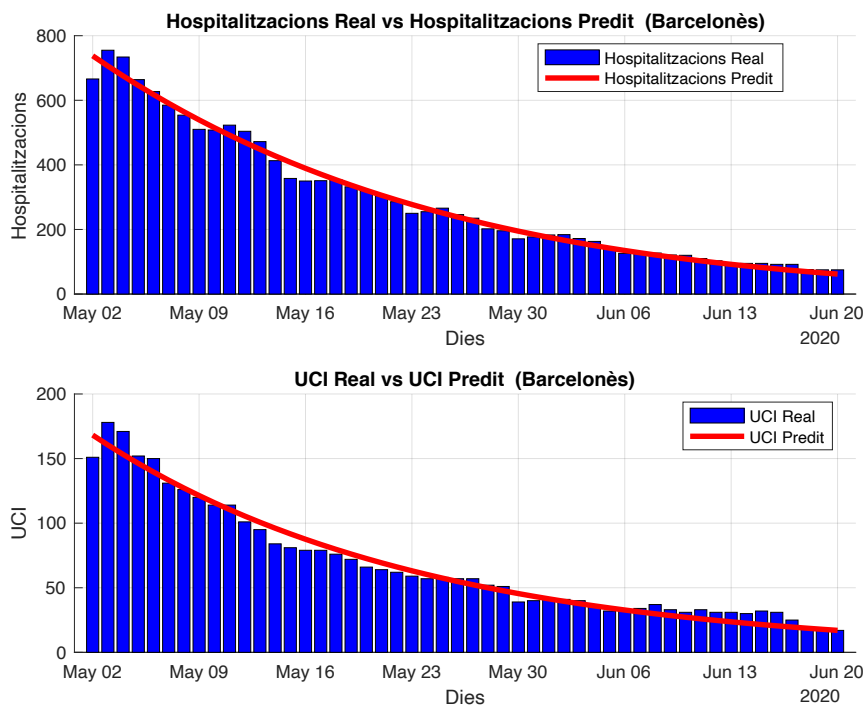
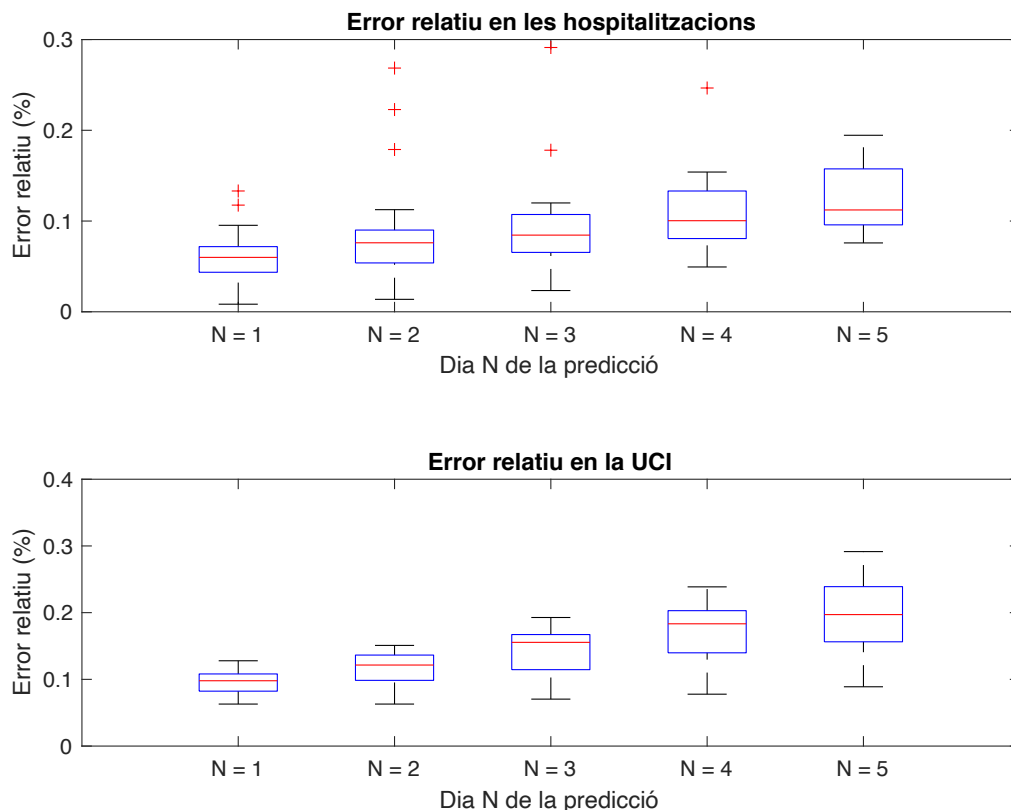


Fig.16. Ajust de la funció de Gompertz en els ingressos a l'UCI i les hospitalitzacions al Barcelonès el 20 de juny de 2020.

	$R^2$	RMSE	K	r	$N_0$
Hospitalitzacions	0,9872	22,42	1636922	-0,0057	770,7
UCI	0,9748	6,91	479881	0,0469	476104,9

Taula VI. Estimació dels paràmetres el 20 de juny de 2020

El model de Gompertz descriu correctament la tendència de les hospitalitzacions i UCI diàries com es veu en la Fig. 16. Els resultats obtinguts presenten una  $R^2 > 0,98$  en les hospitalitzacions i una  $R^2 > 0,97$  en l'UCI. En la Taula VI es mostren els paràmetres ajustats del model.



**Fig.17.** Error relatiu de les prediccions. Anàlisi de la precisió per a prediccions de fins a 5 dies.

S'obté uns errors relatius al voltant del 5% en les prediccions realitzades per 1 i 2 dies en les hospitalitzacions. A mesura que s'augmenta els dies, l'error augmenta fins a un 15%. En el cas de l'UCI, l'error al primer dia és inferior al 10%. Presenta un error relatiu significativament elevat, en prediccions a 5 dies és del 20%. Utilitzant el mateix criteri que en els casos acumulats, es calcula la probabilitat que el valor real pertanyi en l'interval de la predicció. En el cas de les hospitalitzacions s'obtenen uns valors al voltant del 95% pel primer dia, el valor disminueix a mesura que els dies avancen fins arribar a un 75%. En les UCI, per les prediccions del primer dia s'obté un 87% i fins al tercer dia es manté per sobre del 75%. Quan

N=4/5 la probabilitat disminueix considerablement fins a valors del 60%.

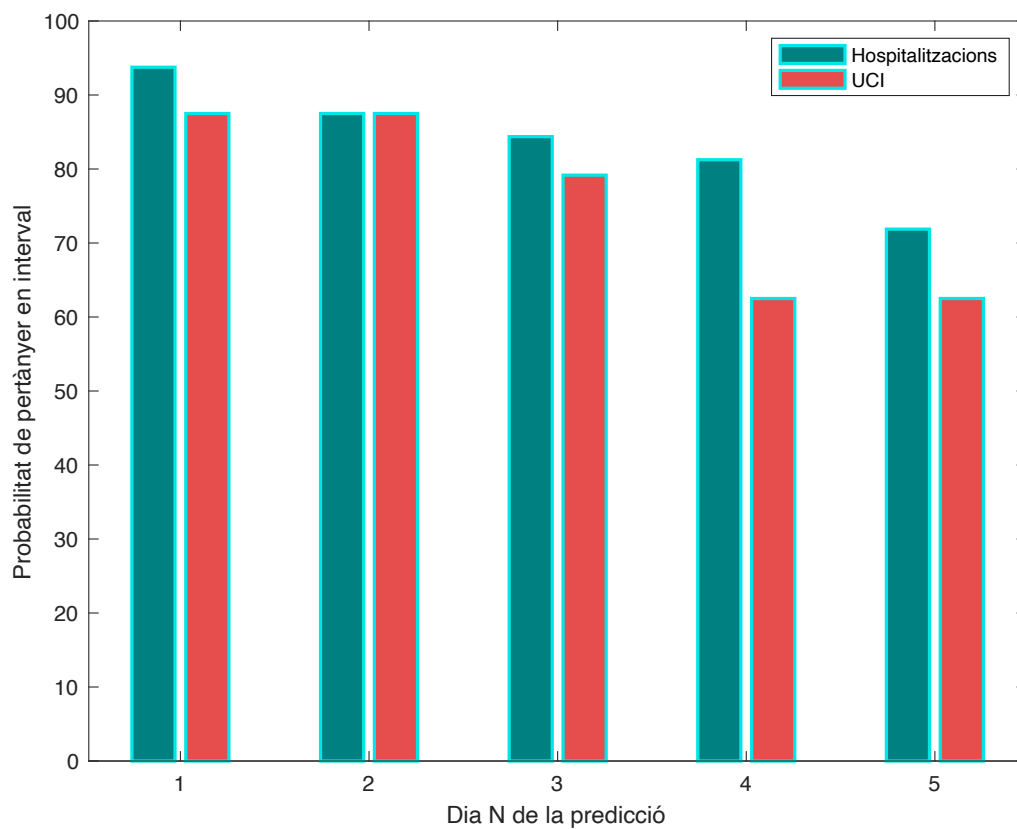


Fig.18. Probabilitat que les prediccions realitzades cauen en l'interval de predicció per cada dia.

## 6. Discussió

El model utilitzat només utilitza les dades sobre els contagis i les morts, que després serveix per ajustar la funció sense tenir en compte altres factors. Es tracta d'una funció empírica, no depèn d'un coneixement previ, és per això que en el cas de la COVID-19 s'ajusta correctament. Tot i que els resultats obtinguts són correctes, existeixen paràmetres que es podrien millorar. Com a qualsevol model, la qualitat de les dades públiques és un factor clau per aconseguir uns resultats satisfactoris. Com que no s'introdueixen altres variables té l'avantatge d'evitar el soroll que provocaria, per contra el model depèn pràcticament de la qualitat de les dades.

En l'estudi no s'inclouen algunes consideracions, com el percentatge de casos asimptomàtics, la cobertura dels tests i el retard en l'obtenció dels resultats del test. És per això, que en futurs estudis es podria incloure-les. Altres limitacions en l'estudi són els casos importats en la regió, així com les mesures de contenció i control aplicades per les institucions. El model no és capaç d'estudiar la reacció en el comportament del virus segons les mesures de control aplicades. Si s'afegeixen els casos asimptomàtics, podrien conduir a brots inesperats a causa de manera desconeguda. Aquests factors poden provocar comportaments desconeguts en la propagació del virus. Un altre obstacle a superar és el resultat del cap de setmana, on el nombre de casos, defuncions i hospitalitzacions varia de manera substancial respecte els altres dies. Això provoca que en les prediccions a curt termini, l'error en les prediccions sigui més elevada. Han aparegut altres problemes com les modificacions en el criteri de reportar les dades, en la Fig.1 s'observa s'ha corregit les dades històriques.

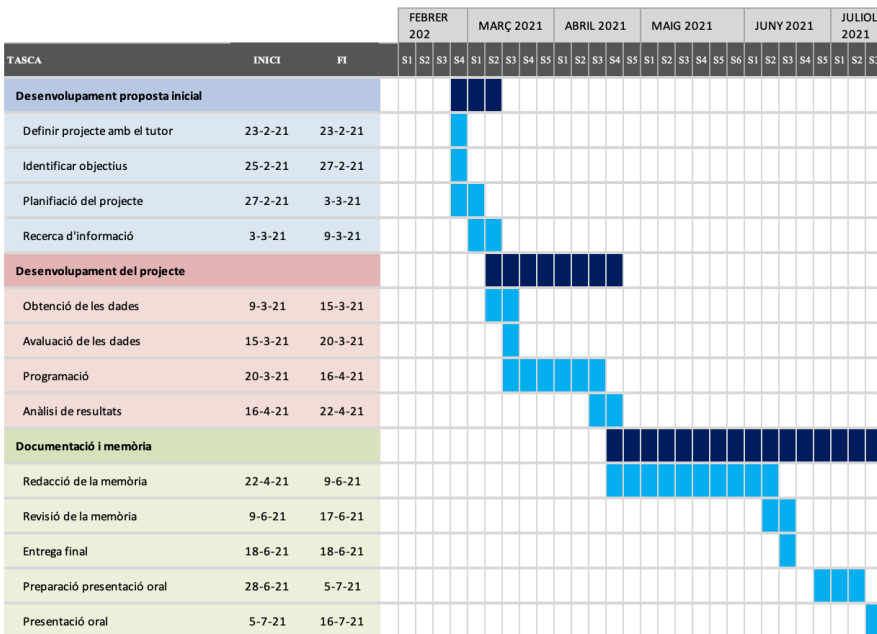
## 7. Planificació temporal

Per aconseguir els objectius proposats en un projecte en un temps limitat, la planificació temporal de les activitats que es realitzaran és un bon mètode. Per organitzar-ho, es divideix el projecte en fases, i dins d'aquestes fases s'hi tenen les activitats a realitzar. Dins de cada fase, també es mostra la duració aproximada de cada activitat i el total d'elles. Per definir la planificació temporal s'utilitza el diagrama de Gantt, és una de les tècniques més comunes en gestió de la qualitat. Presenta les activitats a realitzar durant el projecte, i el calendari complet amb els dies concrets per cada tasca.

La duració del projecte és d'uns tres mesos i està dividit en tres grups. Primer s'hi troba el desenvolupament de la proposició inicial, on s'identifiquen els objectius i abast del projecte i la planificació del projecte. A continuació, està el desenvolupament del projecte, on es realitza la programació del model i anàlisi dels resultats obtinguts. Finalment, l'últim grup consisteix en la documentació i memòria. Dins s'hi troba la redacció i correcció de la memòria, la preparació de la defensa oral i la presentació oral.

### 7.1. Diagrama de Gantt

A continuació es mostra el diagrama de Gantt del Treball de Fi de Grau:





## 8. Estudi econòmic

Els costos d'enginyeria tenen en compte totes aquelles hores invertides a dissenyar el programa de Matlab, així com tots els càlculs realitzats posteriorment. El projecte es divideix en quatre, en la Taula VII es recullen les hores treballades segons els subgrups.

<b>Paràmetre</b>	<b>Temps dedicat [h]</b>
<i>Estudi previ</i>	15
<i>Documentació prèvia</i>	35
<i>Avantprojecte</i>	120
<i>Memòria</i>	80
<i>Total</i>	250

*Taula VII. Hores dedicades a la realització del treball*

El projecte ha estat elaborat per un estudiant d'Enginyeria Industrial. El guany brut mitjà per hora d'un tècnic és de 20,45 € (26), per tant, el cost d'enginyeria ascendeix fins als 5110 €. El programari (Matlab, Office) s'ha utilitzat la seva versió d'estudiant, per tant, el cost ha sigut nul. Tot el treball, així com totes les reunions, s'han realitzat des de casa. Tampoc s'ha de sumar cost de transport ni de lloguer, ja que s'ha realitzat tot des de casa.

## Conclusions

L'objectiu principal ha consistit a estudiar l'evolució i propagació de la COVID-19 a Catalunya mitjançant la funció de Gompertz, i realitzar prediccions a curt i llarg termini a partir d'ajustar les dades reportades per les institucions. La funció de Gompertz sembla un model adequat per l'ús en brots sobtats i universals.

A partir de l'ajust de la funció, s'ha calculat altres paràmetres per entendre el comportament del virus, entre aquests està el pic de casos diaris, el nombre de casos acumulats, la incidència acumulada, etc. Els resultats del model presenta un bon ajust entre els valors reportats i estimats. S'obté una  $R^2 = 0,9984$  en el nombre de casos acumulats i una  $R^2 = 0,9999$  en el nombre de morts acumulades. Com es mostra en els apartats anteriors, els resultats presenten errors relatius inferiors al 3% en les prediccions pel següent dia i de <10% a cinc dies vista. A més, s'obté prediccions superiors on el 85% de casos es troben dins l'interval de predicció. Alhora de l'estudi a llarg termini, s'ha demostrat que el model és vàlid, per exemple, a partir dels paràmetres  $r$ ,  $t_{pic}$ ,  $t_{90}$  i  $K$ . Els resultats presenten un potencial model matemàtic per l'ús en les institucions. A més, s'ha demostrat que el model és vàlid per estudiar les defuncions a partir de les persones amb símptomes. Amb els resultats obtinguts, es pot predir el final de la pandèmia si les mesures de control segueixen adoptades i no apareixen brots desconeguts.

El model permet obtenir l'evolució del nombre d'hospitalitzacions, i mitjançant l'ajust dels paràmetres, permet realitzar prediccions en la capacitat dels hospitals. Tanmateix, el model no s'ajusta prou bé en les UCI, provocant que les prediccions tinguin un error més elevat. Es podrien trobar variacions del model per reduir l'error, però aquest model és capaç d'obtenir resultats que s'ajustin als valors reals.

A tall de recapitulació, el model de Gompertz ha demostrat ser una eina matemàtica capaç de realitzar prediccions a curt termini en l'evolució de la COVID-19, amb la capacitat de col·laborar amb les institucions sanitàries.

## Bibliografia

### Referències bibliogràfiques

1. *Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models.* **Torrealba-Rodriguez, O, Conde-Guitèrrez, RA i Hernández-Javier, AL.** s.l. : Elsevier, 2020, Chaos, Solitons & Fractals, Vol. 138, p. 1009946.
2. *Modeling and forecasting trend of COVID-19 epidemic in Iran until May 13, 2020.* **Ahmadi, Ali, et al.** s.l. : Iran University of Medical Science, 2020, Medical Journal of the Islamic Republic of Iran, Vol. 34, p. 27.
3. *A Bayesian Model of COVID-19 Cases Based on the Gompertz Curve.* **Berihuete, Ángel, Sánchez-Sánchez, Marta i Suárez-Llorens, Alfonso.** s.l. : Multidisciplinary Digital Publishing Institute, 2021, Mathematics, Vol. 9, p. 228.
4. *Framing the Origins of COVID-19.* **Bolsen, Toby, Palm, Risa i Kingsland, Justin T.** s.l. : SAGE Publications Sage CA: Los Angeles, CA, 2020, Science Communication, Vol. 42, p. 562--585.
5. *Review of the clinical characteristics of coronavirus disease 2019 (COVID-19).* **Jiang, Fang, et al.** s.l. : Springer, 2020, Journal of general internal medicine, Vol. 35, p. 1545-1549.
6. *The COVID-19 epidemic.* **Velavan, Thirumalaisamy P i Meyer, Christian G.** s.l. : Wiley-Blackwell, 2020, Tropical medicine & international health, Vol. 25, p. 278.
7. *Estimation of growth parameters using a nonlinear mixed Gompertz model.* **Wang, Z i Zuidhof, MJ.** s.l. : Elsevier, 2004, Poultry science, Vol. 83, p. 847--852.
8. *Prediction and analysis of Coronavirus Disease 2019.* **Jia, Lin, et al.** s.l. : arXiv preprint arXiv:2003.05447, 2020.
9. *The Gompertz curve as a growth curve.* **Winsor, Charles P.** s.l. : National Academy of Sciences, 1932, Proceedings of the National Academy of Sciences of the United States of America, Vol. 18, p. 1.
10. *Modeling of the bacterial growth curve.* **Zwietering, MH, et al.** s.l. : Am Soc Microbiol, 1990, Applied and environmental microbiology, Vol. 56, p. 1875--1881.
11. *The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family.* **Tjørve, Kathleen M. C. i Tjørve, Even.** s.l. : Public

Library of Science San Francisco, CA USA, 2017, PloS one, Vol. 12, p. e0178691.

12. *A unified approach to the Richards-model family for use in growth analyses: why we need only two model forms.* **Tjørve, Even i Tjørve, Kathleen MC.** s.l. : Elsevier, 2010, Journal of theoretical biology, Vol. 267, p. 417-425.

13. *Analysis of growth in intersubspecific crossing of mice using Gompertz model.* **Kurnianto, E, Shinjo, A i Suga, D.** s.l. : Asian Australasian Association of Animal Production Societies, 1998, Asian-Australasian Journal of Animal Sciences, Vol. 11, p. 84-88.

14. *A comparison between Nonlinear Least Squares and Maximum Likelihood estimation for the prediction of tumor growth on experimental data of human and rat origin.* **Patmanidis, Spyridon, et al.** s.l. : Elsevier, 2019, Biomedical Signal Processing and Control, Vol. 54, p. 101639.

15. *Deciphering death: a commentary on Gompertz (1825)'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies'.* **Kirkwood, Thomas BL.** s.l. : The Royal Society, 2015, Philosophical Transactions of the Royal Society B: Biological Sciences, Vol. 370, p. 20140379.

16. *Predicting the number of total COVID-19 cases and deaths in Brazil by the Gompertz model.* **Valle, Jemy A Mandujano.** s.l. : Springer, 2020, Nonlinear Dynamics, Vol. 102, p. 2951–2957.

17. Johns Hopkins University & Medicine. Coronavirus Resource Center. [En línia] [Data: 19 / 4 / 2021.] [https://github.com/CSSEGISandData/COVID-19/tree/master/csse\\_covid\\_19\\_data](https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data).

18. *Comparison of an artificial neural network and Gompertz model for predicting the dynamics of deaths from COVID-19 in México.* **Conde-Gutiérrez, RA, Colorado, D i Hernández-Bautista, SL.** s.l. : Springer, 2021, Nonlinear Dynamics, p. 1-15.

19. *First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment.* **Leung, Kathy, et al.** s.l. : Elsevier, 2020, The Lancet, Vol. 395, p. 1382-1393.

20. *Empirical model for short-time prediction of COVID-19 spreading.* **Català, Martí, et al.** s.l. : Public Library of Science San Francisco, CA USA, 2020, PLoS computational biology, Vol. 16, p. e1008431.

21. Departament de Salut de la Generalitat de Catalunya. [En línia] <https://dadescovid.cat/>.

22. *Health care system dynamics modelization for covid-19.* **Dot Casadevall, Laura i Kirchner Sala, Helena.** s.l. : Universitat Politècnica de Catalunya, 2020.

23. *Robust estimation of diagnostic rate and real incidence of COVID-19 for European policymakers.* **Català, Martí, et al.** s.l. : Public Library of Science San Francisco, CA USA, 2021, PloS one, Vol. 16, p. e0243701.

24. *The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application.* **Lauer, Stephen A, et al.** s.l. : American College of Physicians, 2020, Annals of internal medicine, Vol. 172, p. 577-582. 9.

25. *A Review of Coronavirus Disease-2019 (COVID-19).* **Singhal, Tanu.** s.l. : Springer, 2020, The Indian Journal of Pediatrics, Vol. 87, p. 281-286.

26. Institut d'Estadística de Catalunya. [En línia]  
<http://www.idescat.cat/indicadors/?id=anuals&n=10403>.

## Índex de taules

Taula I. Estimació dels paràmetres segons les regions. ....	17
Taula II. Estimació dels paràmetres el 17 de maig de 2020 .....	21
Taula III. Prediccions dels casos acumulats a Catalunya dels següents 5 dies. ....	22
Taula IV. Prediccions de les morts acumulades a Catalunya dels següents 5 dies. ....	22
Taula V. Comparació de la situació real amb la predicció. ....	29
Taula VI. Estimació dels paràmetres el 20 de juny de 2020 .....	36
Taula VII. Hores dedicades a la realització del treball.....	41

## Índex d'equacions

Equació 1.....	8
Equació 2.....	8
Equació 3.....	9
Equació 4.....	10
Equació 5.....	10
Equació 6.....	10
Equació 7.....	11
Equació 8.....	11
Equació 9.....	11
Equació 10 .....	11
Equació 11 .....	14
Equació 12 .....	28
Equació 13 .....	30

Equació 14.....	32
-----------------	----

## Índex de figures

Figura 1. Nombre total de casos a Hubei el 12 de març de 2020.....	15
Figura 2. Nombre total de casos a Gunagdong i Henan .....	16
Figura 3. Nombre de casos diaris segons la regió.....	17
Figura 4. Ajust dels casos diaris a Catalunya .....	20
Figura 5. Ajust dels casos acumulats a Catalunya .....	20
Figura 6. Error relatiu de les prediccions.....	23
Figura 7. Probabilitat que les prediccions realitzades cauen en l'interval de predicció.....	24
Figura 8. Ajust de la funció de Gompertz a Catalunya segons la data .....	25
Figura 9. Evolució de les prediccions a llarg termini.....	27
Figura 11. Convergència de la predicció del $t_{pic}$ .....	29
Figura 12. $R_t$ a Catalunya .....	31
Figura 13. Evolució del risc de rebrot a Catalunya .....	33
Figura 14. Evolució de la incidència acumulada dels darrers 14 dies .....	33
Figura 15. Comparació entre la IA14 i el nombre de casos diaris.....	34
Figura 16. Ajust de la funció en els ingressos a l'UCI i les hospitalitzacions .....	36
Figura 17. Error relatiu de les prediccions.....	37
Figura 18. Probabilitat que les prediccions realitzades cauen en l'interval de predicció.....	38

## Annex

### Codi de Matlab per ajustar el model de Gompertz a les dades reportades

```
% Importar les dades d'un fitxer .csv

csv = importdata('dades.csv');

Dies = csv.data(:,1);

CasosReals = csv.data(:,2)';

% Definir el model de Gompertz i les condicions inicials

f = @(F,t) F(1)*exp(-F(2)*exp(-F(3)*t));

beta0 = [1000 1 0.2];

mdl = fitnlm(Dies,CasosReals,f,beta0);

% Plot

figure

hold on

plot(Dies,CasosReals,'blue','LineWidth',1.5)

plot(Dies,predict(mdl,Dies),'--','LineWidth',1.5)

xlabel('Data')

ylabel('Casos Acumulats')

title('Casos Reals vs Casos Predits')

legend('Casos Reals','Casos Predits','Location','best')
```



### Predicció dels següents dies

```
Xnew = [42;43;44];  
  
[ynew,ynewci] = predict mdl,Xnew  
  
plotSlice mdl
```

### Càlcul d'altres paràmetres

```
a = mdl.Coefficients.Estimate(3);  
  
KNo = mdl.Coefficients.Estimate(2);  
  
mu0 = a*KNo;  
  
t90 = (-1/a)*log((-log(0.9))/KNo);  
  
tp = (1/a)*log(mu0/a);
```