

# Modelización empírica de la epidemia COVID-19

Marta Ginovart Gisbert  
Mónica Blanco Abellán

**SUMA+** núm. 94  
pp. 11-22

Artículo encargado por *Suma* en abril de 2020 recibido en junio de 2020 y aceptado en julio de 2020

Una epidemia, aparición y propagación de una enfermedad infecciosa en una comunidad, es un proceso dinámico y susceptible de ser modelizado o representado mediante estrategias o perspectivas diversas. El objetivo de este trabajo es presentar, a través de actividades diseñadas e implementadas en clases no presenciales en grados de Ingeniería de Biosistemas y ajustadas al ámbito de los modelos empíricos, cómo se puede empezar a investigar el comportamiento de la epidemia COVID-19 a partir de un conjunto de datos publicados que van desde su inicio en China (mediados de enero de 2020) hasta su llegada a Europa (principios de marzo de 2020).

**Palabras clave:** Datos COVID-19, China, Modelo empírico, Modelo continuo, Programas informáticos.

En estos momentos en que estamos inmersos en una crisis sanitaria sin precedentes por la epidemia COVID-19 (enfermedad contagiosa causada por el coronavirus SARS-CoV-2), la presencia de las matemáticas en los diferentes medios de comunicación es constante. Nos presentan gráficos, tablas de cifras y porcentajes, nos dibujan curvas y nos hablan de su forma, y nos explican la existencia y desarrollo de diferentes tipos de modelos para hacer estimaciones de parámetros epidemiológicos y predicciones futuras.

**Empirical modelling of the COVID-19 epidemic** // An epidemic, considered as the appearance and spread of an infectious disease in a community, is a dynamic process that can be modelled or represented through different strategies or perspectives. The aim of this paper is to present a collection of teaching activities for undergraduates in Biosystems Engineering, focused on empirical models, which were designed to help students investigate the behaviour of the COVID-19 epidemic, from its outbreak in China (mid-January 2020) to its spread to Europe (early March 2020).

**Keywords:** COVID-19 data, China, Empirical model, Continuous model, Software.

Nunca hasta ahora la sociedad en general se había interesado tanto, y de forma tan generalizada, por temáticas y herramientas tan propias de la matemática aplicada.

Una epidemia, con la aparición de un brote y posterior contagio infeccioso en una comunidad, es un proceso dinámico, y por tanto susceptible de ser modelizado con estrategias o perspectivas diversas. La modelización científica tiene el propósito de encon-

trar, interpretar y validar las representaciones aproximadas de estos sistemas, definidos por conjuntos de elementos y conceptos, que tienen características y mecanismos de relación descritos mediante objetos y operaciones matemáticas. Un modelo constituye, pues, una representación abstracta de un cierto aspecto de la realidad. Por ejemplo, el estudio de las variaciones temporales de tamaño y composición de poblaciones biológicas, con un uso extensivo del lenguaje matemático, encuentra interesantes procesos para investigar en el ámbito de la propagación de enfermedades y la dinámica de poblaciones (Bacaër, 2011; Ma y Xia, 2008).

Hay una gran variedad de tipos de modelización que se clasifican de acuerdo a diferentes criterios. La decisión de qué tipo de modelo utilizar para investigar, representar y simular un sistema depende de los objetivos específicos que se persigan en el estudio, y del tipo de información de que se disponga.

Por ejemplo, existen los modelos estáticos que no incluyen el tiempo como una variable y se considera entonces una *foto* o *instantánea* de la realidad a estudiar. En contra, existen los modelos dinámicos, que sí consideran los cambios producidos en el sistema con el paso del tiempo, como manifiestamente es el caso del proceso de propagación de esta epidemia COVID-19. El número de personas infectadas, de personas recuperadas, y de personas fallecidas son algunos de los elementos claves a considerar en la representación de la epidemia y que cambian con el paso del tiempo.

Otra posible componente en la modelización es el espacio (dominio bidimensional o tridimensional) que se puede tener en consideración para la representación aproximada del sistema, o que se puede obviar, dando lugar a modelos espaciales, o no espaciales, respectivamente. Mayoritariamente, los modelos que se presentan en el ámbito académico son modelos no espaciales, ya que el control explícito de coordenadas de posición en un dominio incrementa la complejidad de su diseño, así como su implementación y resolución. Ante la indiscutible importancia del contacto físico o la proximidad entre personas para la propagación del virus SARS-CoV-2, el control ex-

plícito del espacio sería un elemento significativo y de efecto contrastado en el estudio de su propagación. Medidas de confinamiento o restricción de movimiento de las personas en áreas establecidas, así como la delimitación de zonas con características epidemiológicas distintas, podrían ser introducidas en estos modelos espaciales.

---

La decisión de qué tipo de modelo utilizar para investigar, representar y simular un sistema depende de los objetivos específicos que se persigan en el estudio, y del tipo de información de que se disponga.

---

Los modelos cuantitativos o numéricos generalmente incluyen fórmulas y algoritmos matemáticos más o menos complejos, y son los modelos que mayoritariamente se trabajan en el ámbito académico. Se pueden fundamentar en ecuaciones lineales o no lineales, en ecuaciones diferenciales, en el cálculo probabilístico o en métodos estadísticos. Se generan modelos continuos si las variables para representar el tiempo y/o espacio son variables reales, mientras que en los modelos discretos estas son enteras. Por otra parte, existen los modelos cualitativos o conceptuales, que en general predicen si el estado del sistema irá en una determinada dirección, o si aumentará o disminuirá alguna variable (sin importar exactamente la magnitud concreta con la que cambia la variable). En este caso, su interés se centra en reproducir patrones, comportamientos, o tendencias, y son adecuados para explorar distintos escenarios y diferentes respuestas del sistema.

Según la información de entrada, generamos modelos empíricos cuando se utilizan las observaciones directas o los resultados de experimentos del fenómeno que se estudia. En este artículo se trabaja con modelos empíricos a partir de los datos publicados al principio de la epidemia COVID-19 en China. Como estos modelos se nutren de datos, éstos deben ser consistentes para que las estimaciones de los parámetros in-

volucrados en las funciones matemáticas que se manejan sean lo más fiables posible. Hay que conocer las limitaciones propias de los datos disponibles para poder hacer un buen uso de estos modelos, pues la recogida y definición de estos datos condiciona de forma significativa los resultados que se consiguen. Frente a estos modelos empíricos se encuentran los modelos mecanicistas o modelos heurísticos que se fundamentan en las explicaciones sobre las causas o mecanismos naturales que dan lugar al fenómeno que se estudia. Los modelos compartimentales epidemiológicos son de este tipo, consideran que los individuos de una población se encuentran en diferentes compartimentos o estados frente a la enfermedad, como por ejemplo, en estado sano y susceptible al contagio, en periodo de incubación, infectado asintomático, infectado sintomático, recuperado de la enfermedad e inmune (de forma definitiva o únicamente temporal), o fallecido. Los parámetros específicos de la epidemia regulan las transiciones de individuos entre distintos compartimentos o estados. El modelo SIR, por las iniciales en inglés de *Susceptible - Infected* o *Infectious - Recovered* (o *Removed*, pues puede recoger además de los recuperados a los fallecidos), y cuyo origen se sitúa a principios del siglo XX en un trabajo de Kermack y McKendrick en 1927, es el modelo básico sobre el que se han ido desarrollando múltiples variaciones (Hethcote, 2000).

Según la presencia o ausencia de aleatoriedad, se pueden tener modelos deterministas (los que mayoritariamente se trabajan en el ámbito académico) o modelos estocásticos. Los modelos deterministas no utilizan variables aleatorias y se conocen los posibles resultados de forma exacta, no hay incertidumbre en su consecución, y son determinados por el estado inicial. Los modelos estocásticos o probabilísticos introducen el azar a través de componentes aleatorias, por lo que sus resultados no se pueden conocer con exactitud. Es evidente que el proceso de propagación del virus es un proceso estocástico gobernado por múltiples, y a veces desconocidos, factores.

En las últimas décadas han aparecido nuevos tipos de modelización con una amplia base computacional, los modelos computacionales, los cuales requieren de

herramientas y habilidades diferentes a las que se ejercitan trabajando con otras perspectivas de modelización más presentes en los currículos actuales de matemáticas. Por ejemplo, los modelos basados en agentes son modelos computacionales que tratan con sistemas complejos, generando la simulación de acciones e interacciones de elementos autónomos y discretos (agentes) dentro de un determinado entorno (Wilensky y Rand, 2015). Para ilustrar este tipo de modelo se invita al lector a consultar el simulador Virus de la plataforma multiagente NetLogo de acceso libre <<https://ccl.northwestern.edu/netlogo/models/Virus>> y con el que se puede investigar la dinámica de una población virtual en la que se produce la propagación de un virus con individuos sanos, enfermos o contagiados, y recuperados con inmunidad temporal al virus.

Finalmente, según su propósito, los modelos pueden ser predictivos (informan del comportamiento de sus parámetros o valores respuesta en un futuro), evaluativos (prueban diferentes alternativas para poder comparar sus resultados), de optimización (tratan de identificar un máximo o mínimo, o la mejor elección posible), o explicativos (comprueban cómo se conectan las variables para buscar esclarecimientos del fenómeno y contrastar hipótesis formuladas).

La modelización de una epidemia ofrece muchas y atractivas posibilidades para trabajar en clase. En particular, el estudio de la expansión del COVID-19 es una oportunidad para integrar la modelización en los sistemas de enseñanza (Gómez, 2003; Barquero y otros, 2011; Ginovart, 2014).

### Actividades COVID-19: docencia no presencial en tiempos de confinamiento

La modelización permite estudiar en el aula una epidemia desde perspectivas distintas, en general, y en particular, la del COVID-19 que se inició en China, y que se extendió al resto del mundo. Es conveniente conjugar la utilización de distintos tipos de modelos sobre un mismo sistema, ya que estos pueden dar respuestas

complementarias y alimentarse mutuamente de la información que cada uno de estos tipos de modelos proporciona. Los parámetros epidemiológicos de este nuevo virus, que se necesitan como entrada para algunos de los tipos de modelos expuestos anteriormente, se van obteniendo con el análisis de datos recopilados durante los últimos meses, y se van conociendo a través de publicaciones de investigación. La sociedad, y en particular la comunidad científica, saben que los valores de estos parámetros pueden variar y reajustarse a medida que se conozcan más detalles de su transmisión e infectividad. Así pues, los modelos empíricos, los que se alimentan exclusivamente de los datos disponibles, son los primeros modelos que pueden aportar luz sobre el comportamiento de esta epidemia.

---

El objetivo de este trabajo es presentar, a través de actividades implementadas en clase basadas en modelos empíricos, cómo se puede empezar a investigar el comportamiento de la epidemia...

---

En el marco de la docencia no presencial llevada a cabo durante el periodo de confinamiento en los meses de marzo, abril y mayo de 2020, en los primeros cursos de los grados en Ingeniería de Biosistemas de la Universitat Politècnica de Catalunya (UPC), se han diseñado y presentado actividades relacionadas con distintos tipos de modelos que se pueden utilizar para explorar el comportamiento de una epidemia en general, y en particular, de la epidemia COVID-19. En las asignaturas Matemáticas 2, de primer curso y común a todos los grados (135 matriculados), y Programación y Resolución de Problemas, de tercer curso del grado de Sistemas Biológicos (36 matriculados), se han trabajado actividades centradas en la modelización de la COVID-19. Se han desarrollado con éxito a través del campus virtual Atenea y con la ayuda de las herramientas que ofrece GSuite for Education que la UPC ha puesto a disposición de profesores y estudiantes para facilitar, durante este período excepcional, la docencia no presencial.

Las distintas actividades desarrolladas han incorporado y compaginado modelos de tipologías diversas, ofreciendo así al alumnado un abanico de perspectivas de modelización para explorar el comportamiento de una epidemia. En estas actividades, y según el curso, se han trabajado diversos tipos modelos, de forma no excluyente: dinámicos, espaciales, no espaciales, cuantitativos (continuos, discretos), empíricos, heurísticos o mecanicistas, estocásticos, deterministas, computacionales (modelización basada en agentes), predictivos y explicativos. Un análisis detallado de todas las tipologías de modelización tratadas en estas actividades durante diversas semanas por los estudiantes, cae fuera del alcance de este artículo. Así pues, nos centraremos en la presentación de las actividades relacionadas con la modelización empírica de la epidemia COVID-19. El objetivo de este trabajo es presentar, a través de actividades implementadas en clase basadas en modelos empíricos, cómo se puede empezar a investigar el comportamiento de la epidemia a partir de un conjunto de datos recogidos desde los inicios de la epidemia en China (mediados de enero de 2020) hasta su llegada a Europa (principios de marzo de 2020).

A pesar de que estas actividades COVID-19 han sido implementadas en aulas universitarias, consideramos que algunas de ellas se podrían adaptar para ser llevadas a otros niveles educativos, como segundo ciclo de secundaria o bachillerato.

### Actividades COVID-19: primeros datos de partida

El propósito de las primeras actividades diseñadas es analizar y modelizar los primeros datos publicados desde el inicio de la epidemia COVID-19 en China, hasta el momento en que la epidemia se empezó a extender al resto de países del mundo y llegó a Europa. Los datos utilizados se pueden consultar en la ampliación de este artículo (Blanco y Ginovart, 2020) que está disponible en la web de la revista <<http://revistasuma.fespm.es/s94-ginovart-y-blanco-web>>, y fueron publicados el día 5 de marzo de 2020 en dos fuentes de información diferentes:

<<https://thewuhanvirus.com/>> (fuente 1) y <<https://systems.jhu.edu/research/public-health/ncov/>> (fuente 2). Corresponden a 50 días, desde el 15 de enero de 2020 (día 1), cuando fue confirmado el brote del nuevo coronavirus y su propagación en la ciudad de Wuhan en China, hasta el 4 de marzo de 2020, cuando se constató que la epidemia se había extendido a muchos otros países, y, en particular, a España. Los datos corresponden a las siguientes variables:

- Fecha y día (desde el 15/1/2020, día 1, hasta el 4/3/2020, día 50)
- Casos totales (fuente 1), casos en China (fuente 2), y casos en otros países, no de China (fuente 2), casos infectados acumulados.
- Muertes totales (fuente 1), fallecidos acumulados.
- Recuperados totales (fuente 1), y Recuperados totales (fuente 2), acumulados.

Los datos se pueden utilizar directamente como cifras absolutas que intentan reflejar una situación real. Por lo tanto, la representación de esta realidad estará ligada a la calidad y fiabilidad de los datos. Sin embargo, la recopilación de estos datos no es trivial. La complejidad de la situación vivida con la expansión de esta epidemia, condiciona de múltiples maneras la recogida y transferencia de los datos, más aún, cuando el proceso bajo estudio es dilatado en el tiempo, y el conocimiento sobre el nuevo virus avanza durante este tiempo. Los datos que se tenían en el momento de generar las actividades para los estudiantes (primeros días de marzo de 2020) son los datos que se analizarán y presentarán en este trabajo. Asumiendo las limitaciones que todo tipo de estudio observacional puede tener, y más aún, las específicas de la situación epidémica actual, estos datos reportados son utilizados para visualizar tendencias e identificar patrones.

## Utilización de modelos lineales y linealizables

Una de las competencias básicas de la asignatura de Matemáticas 2 de primer curso de los grados de Ingeniería de Biosistemas de la UPC, es: «conocimientos básicos sobre el uso y programación de los

ordenadores, sistemas operativos, bases de datos y programas informáticos con aplicación en ingeniería». En consecuencia, en esta asignatura se trabajan diversas opciones que proporciona una hoja de cálculo y su manejo, así como posibles aplicaciones a problemas en ingeniería. Para los estudiantes de la asignatura Programación y Resolución de Problemas de tercer curso ya se asume un conocimiento básico previo de la hoja de cálculo.

En primer lugar, se presenta un conjunto de tareas que pueden ser realizadas con la ayuda de una hoja de cálculo, y en las que, a partir de la utilización de gráficos, cálculos convenientes, ajustes o regresiones lineales, y comparaciones (entre otras posibles opciones que se puedan considerar adecuadas), se propone visualizar y entender el comportamiento que tuvo la epidemia en China, así como reflejar los inicios de la propagación al resto de los países del mundo. Para la implementación de estas tareas no presenciales, se trabajó con la hoja de cálculo Excel. En las primeras semanas de impartición de esta asignatura de matemáticas se llevó a cabo una práctica presencial que sirvió como introducción al uso de esta hoja de cálculo. No obstante, para un desarrollo ágil y autónomo de la actividad se creyó conveniente insistir o ampliar algunos aspectos sobre el manejo del Excel fundamentalmente en relación a diagramas de dispersión y formatos de ejes, y a la herramienta **añadir tendencia** para obtener ajustes de funciones de los datos representados (Blanco y Ginovart, 2020). Los modelos lineal, exponencial y logístico (como un caso particular de función sigmoidea o función tipo S) ya fueron entrenados previamente por los estudiantes en la resolución de ecuaciones diferenciales ordinarias, ilustrando el estudio de la dinámica de poblaciones. El estudio de las progresiones aritméticas y geométricas también puede ilustrar sobre el tipo de crecimiento o decrecimiento de poblaciones.

Se preparó un primer conjunto de 13 tareas —ver los enunciados en Blanco y Ginovart (2020)— con las que se pautaron representaciones gráficas de las evoluciones temporales completas para las distintas variables reportadas, y los correspondientes ajustes a estos datos a través de funciones diversas, para explorar la adecuación, o no, de su comportamiento a modelos

diversos. También se pautó el análisis detallado de algunos subconjuntos de datos, tanto al inicio como al final, de las evoluciones temporales. Se pidió la generación de variables derivadas de las variables originales, como transformaciones logarítmicas, porcentajes de muertos, muertos diarios, o número diario de casos *nuevos* (o número de casos *nuevos* por caso detectado), para complementar la discusión del comportamiento de la epidemia COVID-19 de forma global en el mundo, y dentro y fuera de China. Se indicó también

que se supervisara la concordancia o discrepancia de las dos fuentes de información utilizadas.

La figura 1 muestra un mosaico de ejemplos de gráficos con datos originales y distintas funciones ajustadas. Son algunas de las representaciones que los estudiantes generaron para dar respuesta a las tareas propuestas. Se ha trabajado fundamentalmente con modelos lineales, exponenciales, logarítmicos, y potenciales, a la vez que se han confirmado también otros tipos de comporta-

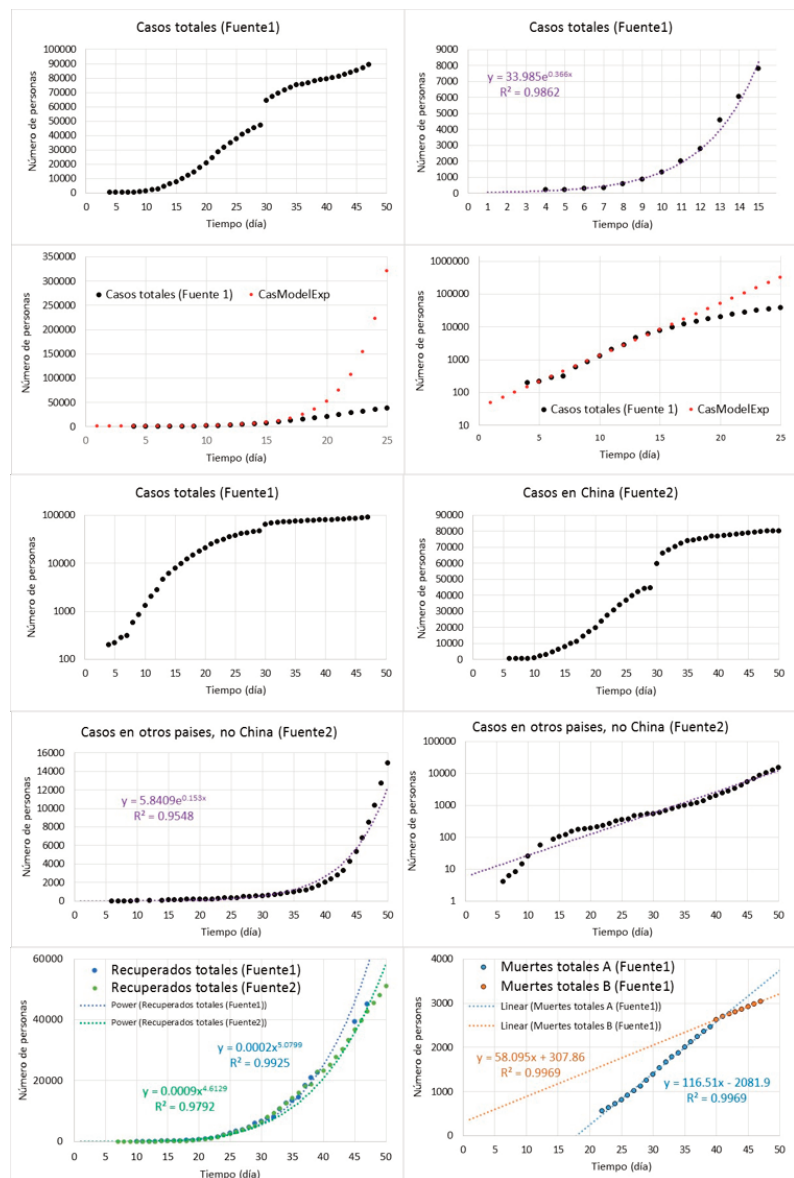


Figura 1. Ejemplos de representaciones gráficas con los datos de las variables originales reportadas de China (Día 1: 15/1/2020, Día 50: 4/3/2020)



mientos, intentando explicar diferentes características de la evolución de la epidemia COVID-19 en China, y en el resto de países del mundo. En algunos casos, cuando los datos son totales, hay una fuerte evidencia de la contribución de los datos únicamente de China, mientras que la contribución que hacen los datos del resto de los países (sin China) en los totales queda condicionada a la magnitud del retraso respecto a la aparición y avance del coronavirus.

La figura 2 muestra otro mosaico de ejemplos de respuestas generadas por los estudiantes en relación a las representaciones gráficas de evoluciones temporales con variables derivadas o generadas a partir de variables

originales. También aquí se puede comprobar la contribución que produce China y la del resto de países en la dinámica de la epidemia COVID-19 durante el periodo temporal analizado. El estudio de las relaciones entre distintas variables también aportó información para comprender el comportamiento de esta epidemia.

Como posible extensión al conjunto de tareas diseñadas, y pensando en alguna actividad opcional y/o actividad dirigida a ciertos estudiantes (los que hubieren mostrado un buen aprovechamiento y un rendimiento alto) se podría complementar este estudio con el uso de medias móviles (herramienta también disponible en la hoja de cálculo). Es un recurso a menudo utilizado en epi-

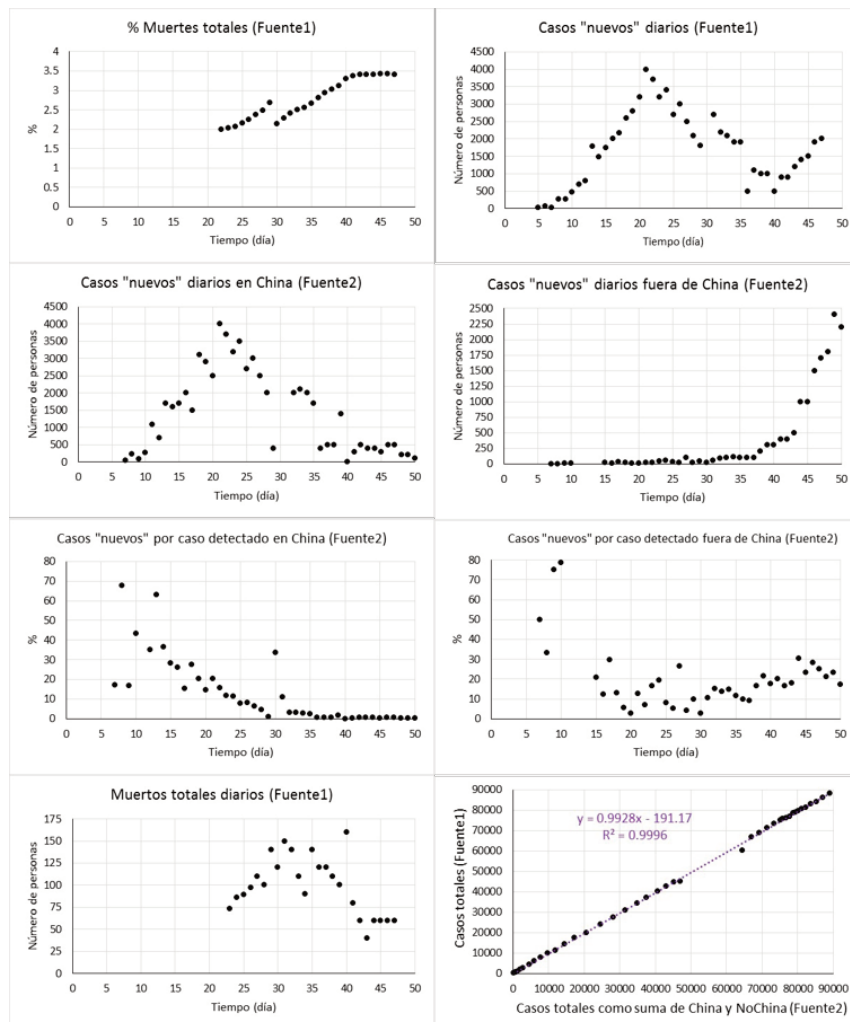


Figura 2. Ejemplos de representaciones gráficas con los datos de las variables derivadas de las originales reportadas de China (Día 1: 15/1/2020, Día 50: 4/3/2020)

demiología para eliminar la variabilidad diaria que puede ser considerable en algunas ocasiones, y suavizar la representación gráfica de las evoluciones temporales.

El cálculo de determinados índices, indicadores o parámetros (que pueden involucrar diversos días anteriores y posteriores a una determinada fecha), y que los epidemiólogos pueden utilizar para estudiar características de la enfermedad, podría generar nuevas tareas que deberían ser acompañadas de una contextualización epidemiológica apropiada, y que a nuestro entender serían más adecuadas para cursos superiores.

## Utilización de modelos no lineales

En una de las tareas anteriores, la forma de la evolución temporal de la variable *casos en China* (fuente 2) se relacionó con la de las funciones sigmoides o funciones tipo S, pero se constató que estas funciones no figuraban entre las opciones que proporciona la hoja de cálculo para efectuar el ajuste de datos. Estas funciones sigmoides modelizan muy bien el comportamiento de la epidemia en China, con un brote infeccioso inicial, un periodo de expansión y un posterior control de esta (fundamentalmente con medidas de confinamiento y restricciones de movilidad). Se produce un crecimiento inicial de casos infectados, con una propagación aproximadamente exponencial, que al cabo de un tiempo se va ralentizando con la disminución de los incrementos diarios de casos infectados (figura 1). Cuando la curva, guiada por los datos, experimenta un cambio de concavidad, la derivada alcanza el valor máximo, y es entonces cuando se evidencia la disminución del ritmo de crecimiento de casos reportados. Esta disminución progresiva de la velocidad de crecimiento hasta valores muy próximos a cero conduce a la curva a alcanzar un valor aproximadamente estable de casos reportados. En la figura 1 se puede observar este comportamiento con la representación de la variable: *casos nuevos* diarios en China (fuente2).

En el marco de la asignatura de Programación y Resolución de Problemas, de tercer curso del grado de Sistemas Biológicos, se llevaron a cabo dos tareas

COVID-19 más, con referencias explícitas al uso de funciones sigmoides (Blanco y Ginovart, 2020). Para la consecución de estos modelos no lineales los estudiantes utilizaron el programa estadístico Minitab que realiza regresiones no lineales. Los estudiantes ya conocían este programa pues fue utilizado previamente en la asignatura de Estadística de segundo curso. Además, Minitab forma parte del conjunto de programas que la UPC distribuye a los estudiantes para su instalación en ordenadores personales.

Minitab permite investigar diferentes tipos de funciones sigmoides, y fueron tres las opciones que se utilizaron para modelizar los datos de la evolución de la epidemia en China, el modelo Gompertz, el modelo logístico, y el modelo loglogístico (figura 3). Estos modelos representan bien la naturaleza de la evolución de la epidemia con la explosión inicial de contagios por el coronavirus durante una primera etapa temporal, seguida de un descenso en la velocidad de propagación del virus durante una segunda etapa, para llegar a una tercera etapa con incrementos de casos prácticamente nulos y una estabilización del número acumulado de casos infectados.

Se utilizó el algoritmo iterativo Gauss-Newton con la entrada de valores iniciales para los parámetros de las funciones, requeridos para poder iniciar las iteraciones. Este algoritmo, a partir de estos valores iniciales, va ajustando sistemáticamente las estimaciones de los parámetros para reducir la suma de los cuadrados del error residual (discrepancias entre valores observados y valores predichos por la función a ajustar). La salida del programa proporciona información sobre el método iterativo, la valoración de la bondad del ajuste conseguido, las estimaciones puntuales y por intervalos de confianza de las estimaciones obtenidas para los parámetros, así como sobre la predicción de futuros valores (<<https://support.minitab.com/en-us/minitab/18/>>).

Las tres funciones sigmoides utilizadas resultaron muy apropiadas para la modelización de la evolución temporal completa de la COVID-19 en China. Una cuestión que se planteó fue la de investigar si estas funciones sigmoides podrían ser también útiles para hacer predicciones sobre la propagación de la epide-



mia aun cuando esta no se hubiera consumado, es decir, aun sin disponer de su evolución temporal completa. Se puede indagar si con un número menor de datos, únicamente con datos de la primera parte de la evolución temporal, y antes de alcanzar la forma característica de la curva con el cambio de concavidad y la estabilización de casos, estas funciones tipo S pueden ser de utilidad. Si se asume que la evolución de la epidemia en una población seguirá este tipo de comportamiento, se podría inspeccionar si ajustar una función tipo S a una primera parte de la evolución temporal, con el brote y la consecuente propagación del virus, el ajuste que se consiga respondería suficientemente bien para avanzar sobre el futuro de esta evolución. Con los datos disponibles, se puede com-

probar si las estimaciones de los parámetros implicados en estas funciones a partir de un subconjunto de datos correspondientes al inicio de la evolución resultan suficientes para definir con acierto la función completa que ya se conoce. También se puede comprobar si la función conseguida con esta evolución parcial tiene capacidad para generar predicciones futuras acertadas respecto a la realidad conocida. Con estas ideas, a primeros de abril de 2020, se expuso este reto a los estudiantes (Blanco y Ginovart, 2020).

Se escogieron para estas regresiones no lineales los primeros datos de la evolución temporal hasta el día 24. Las figuras 4, 5 y 6 recogen los resultados correspondientes al modelo Gompertz, modelo logístico,

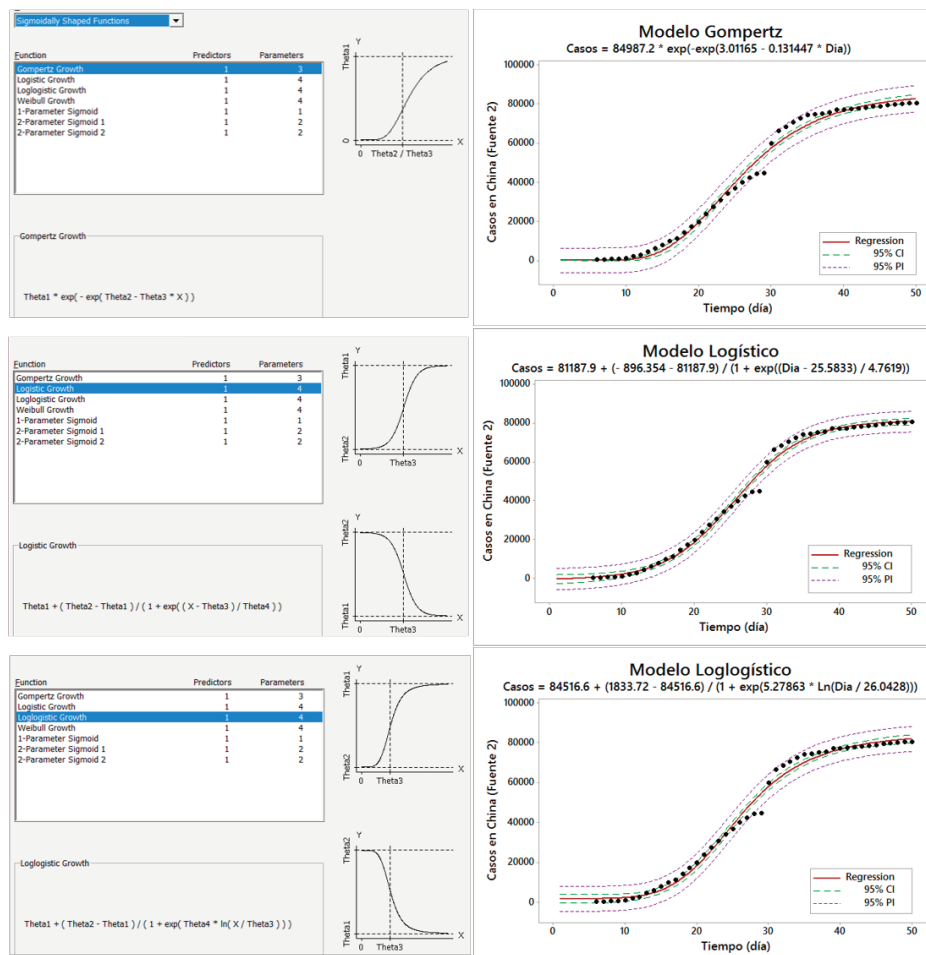


Figura 3. Funciones sigmoideas para modelizar la evolución de casos infectados (acumulados) en China (fuente 2) desde el día 1, 15/1/2020, al día 50, 4/3/2020, con los modelos Gompertz, logístico, y loglogístico

y modelo loglogístico, respectivamente. En estas figuras se muestran las representaciones graficas de estos datos y de las tres funciones ajustadas, las estimaciones puntuales y por intervalos de confianza de los parámetros implicados, y las predicciones de casos futuros para los dos días posteriores a los utilizados en el ajuste, los días 25 y 26. Las estimaciones puntuales de estas predicciones (Fit) se acompañan de los intervalos de confianza (CI) y de predicción (PI) co-

rrespondientes. Estos resultados se comparan con la información conseguida anteriormente con el ajuste del conjunto completo de datos (figura 3) y con los casos reales observados en China para esos dos días (Blanco y Ginovart, 2020). Las tres regresiones realizadas sobre el subconjunto de los primeros datos observados de la epidemia recogen de forma razonable el comportamiento real y observado de los datos completos, siendo el modelo loglogístico el que pro-

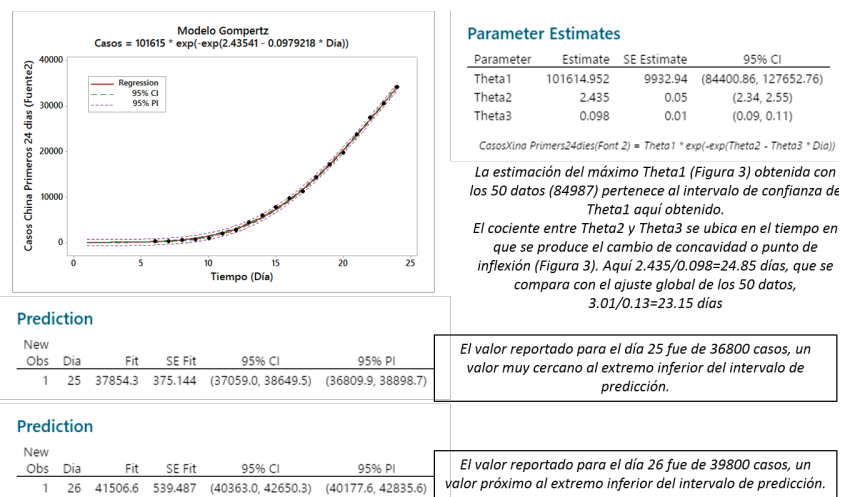


Figura 4. Resultados correspondientes al modelo Gompertz ajustando los datos desde el día 1, (15/1/2020) hasta el día 24 (7/2/2020) de casos (infectados acumulados) en China

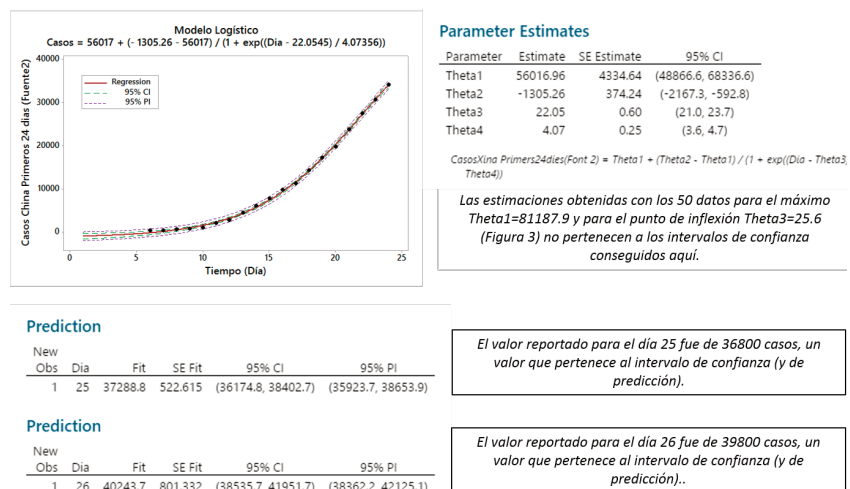


Figura 5. Resultados correspondientes al modelo logístico ajustando los datos desde el día 1, (15/1/2020) hasta el día 24 (7/2/2020) de casos (infectados acumulados) en China

porcionó en general mejores resultados. Este subconjunto formado por los primeros datos de la evolución parece ser suficiente para *forzar* o *perfilar* la forma de las funciones ajustadas. A medida que se utilicen más datos para efectuar las regresiones no lineales y se avance más en la visualización de la forma S, más precisas serán las estimaciones y las predicciones.

Conseguida una modelización satisfactoria para la evolución de casos (infectados) acumulados en China utilizando funciones sigmoides, se podría investigar el uso de estas funciones para otros contextos o países con brotes de COVID-19. Por ejemplo, ¿se podría aprovechar el conocimiento sobre el comportamiento de la epidemia COVID-19 en China para hacer predicciones en otras poblaciones u otros países, con un contexto epidémico similar?

Situados en el momento temporal en que se realizaron estas actividades con los estudiantes, finales de marzo de 2020, y considerando lo que estaba pasando en Cataluña y en España en esas fechas, era razonable y deseable pensar que ese rápido crecimiento inicial de infectados que se vivía, debería ir disminuyendo con el paso del tiempo (como ya había pasado en China), pasando por algún momento temporal (desconocido entonces) que mostraría un punto de in-

flexión, disminuyendo el incremento de casos diarios, hasta llegar a un número de infectados (acumulados) aproximadamente estable. Era previsible que la evolución temporal de la epidemia que vivíamos en Cataluña, y en España, pudiera tener un comportamiento similar al que había tenido en China (con medidas de confinamiento similares). Se confiaba en que se produjera el comportamiento típico de las curvas tipo S para esas evoluciones temporales de casos de COVID-19, aunque, naturalmente, con distintos valores para los parámetros implicados en la función según fuera la zona geográfica bajo estudio (y distintos a los conseguidos en China). Estos valores deberían reflejar características específicas de la dinámica de la población infectada, como el momento temporal en que se produciría el cambio de concavidad, el número máximo acumulado de infectados, o el incremento diario máximo de casos reportados. Con estas reflexiones, se planteó una tarea opcional a los estudiantes para realizar con datos de la epidemia COVID-19 en Cataluña y en España (Blanco y Ginovart, 2020).

Todas las predicciones se deben tomar con mucha prudencia y precaución, ya que conciernen a evoluciones temporales de una epidemia nueva que se estudia a partir de modelos empíricos en un contexto

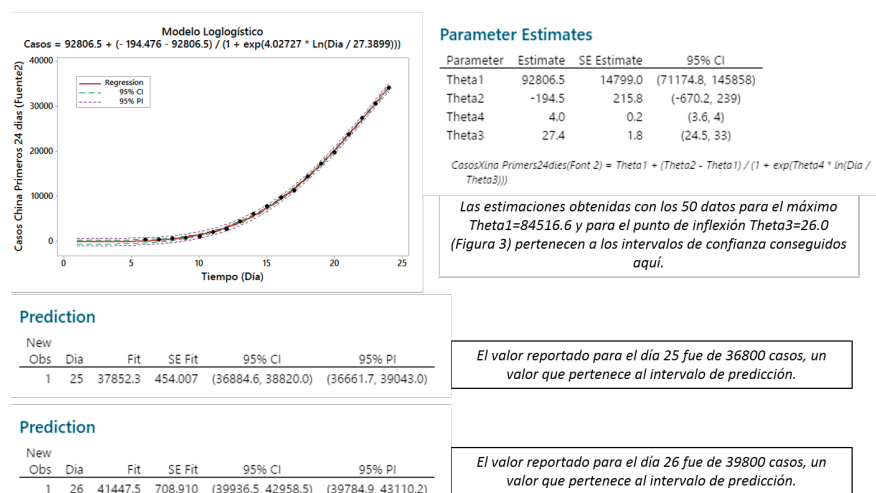


Figura 6. Resultados correspondientes al modelo loglogístico ajustando los datos desde el día 1, (15/1/2020) hasta el día 24 (7/2/2020) de casos (infectados acumulados) en China

académico. Por otra parte, el sistema podría alterar su dinámica de forma súbita e imprevisible, por la acción de factores incontrolados y desconocidos hasta este momento, o por acciones externas sobre el sistema que podrían romper la naturaleza propia de la dinámica de la epidemia en la población.

A pesar de todas las consideraciones comentadas (y otras que se podrían añadir), la modelización empírica es utilizada por grupos de investigación para estudiar el comportamiento de la epidemia COVID-19 en los diversos países del mundo y en zonas geográficas limitadas, y para alimentar con sus salidas o resultados las entradas que necesitan otros tipos de modelos (Blanco y Ginovart, 2020).

Los datos recopilados y publicados de la COVID-19 constituyen una fuente de información de incalculable valor que hay que manejar acertadamente, y es por eso que un adecuado análisis de estos datos proporciona conocimiento valioso sobre su comportamiento. Con el diseño de este conjunto de tareas implementadas en clase, y dirigidas a estudiantes de diversos cursos de los grados de Ingeniería de Biosistemas de la UPC, se ha entrenado el proceso de modelización empírica con datos contextualizados, de un interés innegable, conectando con información que reciben los estudiantes constantemente a través de medios de comunicación. El papel relevante que tienen funciones lineales, exponenciales y sigmoides se ha evidenciado en la descripción y representación del comportamiento de la epidemia COVID-19 en China (y en el resto del mundo).

## Referencias bibliográficas

- BACAËR, N. (2011), *A Short history of mathematical population dynamics*, Springer-Verlag, London.
- BARQUERO, B., M. BOSCH y J. GASCÓN (2011), «Los recorridos de estudio e investigación y la modelización matemática en la enseñanza universitaria de las ciencias experimentales», *Enseñanza de las ciencias*, n.º 29 (3), 339-352.
- BLANCO, M., y M. GINOVART (2020), Ampliación del artículo «Modelización empírica de la epidemia COVID-19», publicado en *Suma*, n.º 94, <<http://revistasuma.fespm.es/s94-ginovart-y-blanco-web>>.
- DIEKMANN, O., H. HEESTERBEEK y T. BRITTON (2013), *Mathematical tools for understanding infectious disease dynamics*, Princeton University Press.
- GINOVART, M. (2014), «Discovering the power of individual-based modelling in teaching and learning: The study of a predator-prey system», *Journal of Science Education and Technology*, 23, 496-513.
- GÓMEZ, J. V. (2003), «La modelización matemática: una herramienta válida en la enseñanza de las matemáticas universitarias», *Suma*, n.º 42, 37-46.
- HETHCOTE, H. W. (2000), «The mathematics of infectious diseases», *SIAM Review*, n.º 42 (4), 599-653.
- MA, S., y Y. XIA (2008), *Mathematical understanding of infectious disease dynamics*, Lecture Notes Series, Volume 16, World Scientific Publishing Co. Pte. Ltd, Singapore.
- WILENSKY, U., y W. RAND (2015), *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*, MIT Press, Hong Kong.

---

### Marta Ginovart Gisbert

Universitat Politècnica de Catalunya  
<[marta.ginovart@upc.edu](mailto:marta.ginovart@upc.edu)>

### Mónica Blanco Abellán

Universitat Politècnica de Catalunya  
<[monica.blanco@upc.edu](mailto:monica.blanco@upc.edu)>