

Modelo epidemiológico de predicción de la COVID-19 en España, Desarrollo de Sistemas Inteligentes

JAVIER CÓRDOBA ROMERO and JUAN JOSÉ CORROTO MARTÍN

En este trabajo introducimos al lector a los modelos matemáticos epidemiológicos y tratamos de extender los modelos clásicos a la predicción de múltiples variables de la COVID-19 dentro de las diferentes comunidades autónomas españolas para que este sea aplicable a los datos proporcionados por las autoridades.

También exploramos la solución numérica del sistema de ecuaciones diferenciales planteado y los problemas inherentes que acarrea la solución del mismo aplicados al lenguaje de programación MATLAB.

La conclusión alcanzada es que el modelo propuesto podría predecir correctamente el desarrollo de la COVID-19 en España con una optimización de parámetros correcta.

Additional Key Words and Phrases: time series, epidemiological models, machine learning, ordinary differential equations

ACM Reference Format:

Javier Córdoba Romero and Juan José Corroto Martín. 2020. Modelo epidemiológico de predicción de la COVID-19 en España, Desarrollo de Sistemas Inteligentes. 1, 1 (May 2020), 5 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 MODELOS EPIDEMIOLÓGICOS

Los modelos epidemiológicos intentan predecir la evolución de una enfermedad infecciosa. El principio de funcionamiento de estos modelos consiste en asignar a los individuos de una población a diferentes compartimentos definidos por el modelo junto con las leyes que gobiernan la transición entre estos compartimentos.

Estos modelos fueron inicialmente propuestos en [Kermack and McKendrick 1927], este modelo inicial se conoce como *modelo SIR*, ya que dividía a la población en tres compartimentos, Susceptibles, Infectados y Recuperados, en este último compartimento se incluyen tanto los fallecidos como los curados de la enfermedad.

En este modelo se realiza una suposición muy relevante respecto al comportamiento de la enfermedad, los recuperados no pueden infectar a nuevos individuos, es decir, el hecho de que un individuo se recupere de la enfermedad impide que este suponga un peligro de contagio para el resto de la población susceptible.

1.1 El modelo SIR

Este modelo básico cuenta con dos parámetros, la tasa de contagio, β y la tasa de recuperación γ , otro valor que se puede extraer del

modelo es el número de reproducción básico o R_0 , este número representa a cuántas personas sanas infecta un infectado, categorizado en tres tipos de evolución [Hethcote 2008]:

- Si $R_0 > 1$: La enfermedad seguirá creciendo progresivamente.
- Si $R_0 = 1$: La enfermedad se estabilizará en el tiempo.
- Si $R_0 < 1$: La enfermedad irá decreciendo poco a poco hasta desaparecer.

Este modelo básico se puede representar mediante el siguiente sistema de ecuaciones diferenciales, ya que este representa cuántas personas se mueven entre los diferentes estados:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta * S * I}{N_c} \\ \frac{dI}{dt} &= \frac{\beta * S * I}{N_c} - \gamma * I \\ \frac{dR}{dt} &= \gamma * I\end{aligned}$$

Donde N_c es la población total. Otro supuesto de este modelo es que el número de nacimientos y muertes naturales es mucho menor al número de infecciones por la enfermedad, si se supone esta condición se cumple la siguiente igualdad:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$$

Esta igualdad da a entender que la población total no cambia con el paso del tiempo.

Si el sistema de ecuaciones diferenciales es resuelto de forma numérica, se obtiene la gráfica de la Figura 1. A este modelo se le denomina modelo SIR sin dinámicas vitales.

1.2 Mejoras respecto al modelo SIR básico

El problema del modelo SIR básico es el hecho de que no refleja las medidas de prevención tomadas para prevenir el contagio, algunos modelos intentan paliar estas limitaciones tomando el parámetro β o γ como una función para simular el efecto de las medidas preventivas aplicadas, como se hizo en [Bjørnstad et al. 2002] aplicado al sarampión.

Con el objetivo de disminuir el error del modelo, a estos modelos se le pueden añadir múltiples compartimentos con sus correspondientes tasas de movimiento, otros modelos más complejos también populares son:

- **SIRD**: Susceptible, Infectado, Recuperado, Fallecido (Deceased)
- **SEIR**: Susceptible, Expuesto, Infectado, Recuperado (Tiene en cuenta el periodo de incubación)
- **SEIS**: Susceptible, Expuesto, Infectado, Susceptible (Sin inmunidad adquirida)

Authors' address: Javier Córdoba Romero, javier.cordoba1@alu.uclm.es; Juan José Corroto Martín, JuanJose.Corroto@alu.uclm.es.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Association for Computing Machinery.

XXXX-XXXX/2020/5-ART \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

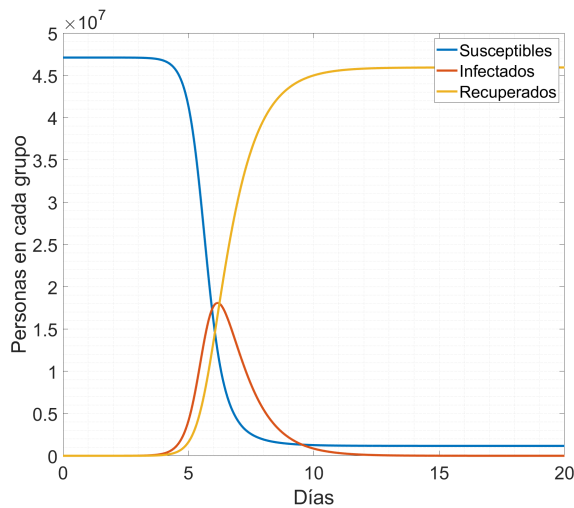


Fig. 1. Gráfico de un modelo SIR básico después de ser resuelto el sistema de ecuaciones diferenciales, con parámetros $\beta = 4$ y $\gamma = 1.057$ y constante de población 47100396 (España)

1.3 Consideraciones a la hora de optimizar el modelo

Como se puede observar en la Figura 1, la curva amarilla representa el número de individuos recuperados acumulados en el tiempo. Sin embargo, en el caso de la curva naranja, los infectados, representa cuántos casos había en cada instante de tiempo.

De este hecho se puede inferir lo siguiente, en este tipo de modelos hay dos compartimentos:

- **Compartimentos intermedios:** Son aquellos por los cuales los individuos pasan un tiempo finito hasta que viajan a otro compartimento del modelo. El valor de estos respecto al tiempo será el número de individuos que había en aquel instante de tiempo en concreto.
- **Compartimentos finales:** Son aquellos a los cuales los individuos van a parar al final de su viaje por los distintos compartimentos del modelo. El valor de estos respecto al tiempo será un acumulado de todos los individuos que se encuentran en este compartimento.

El valor de este tipo de compartimentos tendrá que ser tenido en cuenta a la hora de obtener los errores respecto a los datos reales ya que si se usan datos acumulados frente a datos reales instantáneos la tasa de errores será inexacta.

2 MODELO EPIDEMIOLÓGICO PLANTEADO

En el modelo epidemiológico que planteamos hemos intentado incorporar todas las mejoras propuestas al modelo SIR básico mientras que se añadían otros grupos específicos del comportamiento específico de la enfermedad en España junto con el comportamiento sugerido por el conocimiento que se ha ido obteniendo a partir de las noticias y el conocimiento científico acumulado.

En este modelo no se han tomado en cuenta las dinámicas vitales ya que la epidemia de la COVID-19 ha alcanzado el pico de contagios en menos de dos meses desde que empezó en España, por lo que se

estima que estas dinámicas tendrán poco efecto en el modelado de la enfermedad.

El modelo propuesto está compuesto por 10 compartimentos y 16 parámetros a optimizar, también representado en la Figura 2:

$$\begin{aligned}\frac{dQ}{dt} &= \alpha(t) * S - \frac{\beta_Q * Q * I}{N_c} \\ \frac{dS}{dt} &= -\alpha(t) * S - \frac{\beta * S * I}{N_c} \\ \frac{dL}{dt} &= \frac{\beta_Q * Q * I}{N_c} + \frac{\beta * S * I}{N_c} - \theta_L * L - \kappa_L * L \\ \frac{dI}{dt} &= \theta_L * L - \delta_I * I - \gamma_I * I \\ \frac{dA}{dt} &= \kappa_L * L - \gamma_A * A \\ \frac{dH}{dt} &= \delta_I * I - \gamma_H * H - \tau_H * H - \sigma_H * H \\ \frac{dU}{dt} &= \sigma_H * H - \tau_U * U - \rho_U * U \\ \frac{dUR}{dt} &= \rho_U * U - \gamma_{UR} * UR \\ \frac{dR}{dt} &= \gamma_I * I + \gamma_A * A + \gamma_H * H + \gamma_{UR} * UR \\ \frac{dD}{dt} &= \tau_H * H + \tau_U * U\end{aligned}$$

Los compartimentos son los siguientes:

- **Q (En cuarentena):** Individuos que una vez decretado el estado de alarma se confinaron en sus casas para prevenir la propagación de la enfermedad.
- **S (Susceptibles):** Individuos sanos y susceptibles de ser contagiados por otros Infectados, en situación de cuarentena estos individuos representarán aquellos que presten servicios esenciales y por tanto no puedan confinarse.
- **L (Latentes):** Individuos que han sido contagiados pero aún no son capaces de contagiar ni tampoco muestran síntomas
- **I (Infectados):** Individuos que han sido contagiados, muestran síntomas y son capaces de contagiar.
- **A (Asintomáticos):** Individuos que han sido contagiados, no muestran síntomas, no son capaces de contagiar ni podrán pasar al compartimento de Infectados.
- **H (Hospitalizados):** Individuos infectados que precisan de hospitalización debido a los síntomas de la enfermedad.
- **U (Críticos, en UCI):** Individuos hospitalizados que precisan de ingreso en UCI debido a su estado crítico.
- **UR (Críticos ya recuperados, post UCI):** Individuos que previamente estuvieron ingresados en la UCI pero ya están recuperados y se recuperarán.
- **R (Recuperados):** Individuos que previamente estuvieron infectados y podían contagiar pero actualmente no tienen

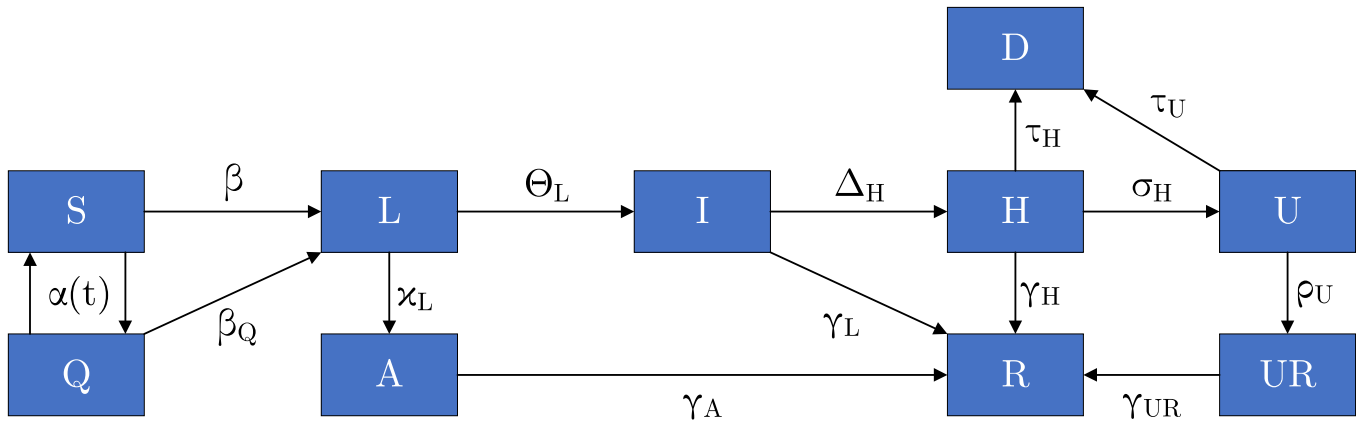


Fig. 2. Diagrama de estados del modelo epidemiológico propuesto.

síntomas, no pueden contagiar y han desarrollado inmunidad a la enfermedad ¹.

- **D (Fallecidos):** Individuos que fallecieron a causa de los síntomas de la enfermedad.

Los parámetros a optimizar más importantes son los siguientes:

- $\alpha(t)$: Porcentaje de individuos en el instante de tiempo t que pasarán al compartimento Q desde el compartimento S, este término podrá ser negativo para simular una desescalada.
- β : Ratio de individuos que pasarán al compartimento L desde S, este parámetro estará dividido en dos parámetros, β_B y β_A . El primero de ellos representará el ratio antes de la cuarentena y el segundo el ratio después de la cuarentena.
- $L(0)$: Número de individuos que estaban en el compartimento L cuando empezó la epidemia.

Este modelo está basado en los siguientes supuestos:

- Una persona enferma de COVID-19 (en compartimento I) no puede fallecer sin antes pasar por el compartimento *hospitalizado* (compartimento H).
- Una persona en estado crítico (ingresada en UCI, compartimento U) y una vez se recupera y baja a planta (compartimento UR) no puede volver a estar ingresada en la UCI y, por tanto, no podrá fallecer.

2.1 Ventajas y Desventajas

Una de las ventajas del modelo es que permite modelar las distintas fases de la desescalada propuestas por el Gobierno haciendo que gente que estaba en el compartimento Q pasar al compartimento S.

Una de las debilidades del modelo es que carece de la capacidad de modelar las interacciones entre las distintas comunidades que conforman España, por lo que un infectado en Madrid no podría infectar a una persona en Huesca.

Esta será una de las cuestiones a tener en cuenta a la hora de resolver el sistema de ecuaciones planteado.

¹La evidencia científica todavía no ha llegado a la conclusión de que los individuos recuperados no puedan contagiar la enfermedad [Organización Mundial de la Salud 2020]

2.2 Posibles mejoras al modelo

Una de las posibles mejoras que planteamos para mejorar la precisión del modelo, incluso antes de intentar optimizar los parámetros, es usar un sistema de ecuaciones diferenciales con retardo en vez de un sistema de ecuaciones diferenciales ordinarias para modelar cómo se comportaría la enfermedad.

Un ejemplo de esto es que, los infectados en el día 15 no dependen de los latentes del día 15, sino de los latentes del día 15 menos el tiempo de incubación de la enfermedad.

3 RESOLUCIÓN EN MATLAB Y PROBLEMAS

MATLAB permite resolver sistemas de ecuaciones diferenciales mediante múltiples métodos dependiendo de distintas características del modelo ², MATLAB recomienda usar ode45 para la mayoría de problemas a resolver por lo que decidimos escoger este método.

Sin embargo, mientras íbamos escalando el modelo desde el SIR básico a otros más complejos nos encontramos con un problema: aún después establecer $\alpha(t) = 1$, es decir, que toda la gente pasase al compartimento Q, en la solución proporcionada por el método nadie pasaba al compartimento Q.

Después de analizar cómo se comportaría el sistema alrededor $t = \text{Día Cuarentena}$, concretamente, después de dibujar cómo deberían ser las gráficas de las personas en los compartimentos Q y S (Figura 3) se encontró el problema.

El problema encontrado era que el sistema es discontinuo en $t = \text{Día Cuarentena}$ con discontinuidades de salto finito en los compartimentos Q y S.

Esto es debido a que el evento que hace que los individuos transicionen hacia el compartimento Q es un evento discreto y no continuo, por lo tanto el método que resuelve el sistema de ecuaciones no "ve" el evento.

En el caso de que se forzase al método a que lo vea, como también hemos intentado para resolver este problema, la derivada que tomaría el método de resolución de sistemas sería inconsistente con el resto de derivadas cercanas al punto y seguiría sin funcionar correctamente.

²<https://www.mathworks.com/help/matlab/math/choose-an-ode-solver.html>

Table 1. Parámetros ficticios para la simulación del comportamiento del modelo

Parámetro	$\alpha(t)$	$L(0)$	β_B	β_A	β_Q	θ_L	κ_L	γ_A	δ_H	γ_I	γ_H	τ_H	σ_H	τ_U	ρ_U	γ_{UR}
Valor	0.65	20	1.3	0.649	0.15	0.72	0.12	0.2	0.4	0.23	0.631	0.37	0.12	0.74	0.7	0.2

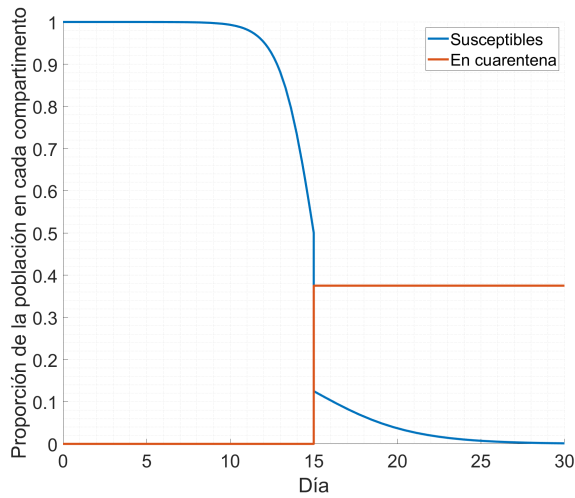


Fig. 3. Gráfico del comportamiento de un modelo QSIR en el cual los individuos transicionan al compartimento Q a través de un evento discreto. Se puede observar como existe una discontinuidad de salto finito en $t = 15$.

La solución al problema es partir la solución en dos intervalos $[t_0, t]$ y $[t, t_f]$ y resolver cada intervalo por separado, las condiciones iniciales del segundo intervalo serán la solución final del primer intervalo y las transformaciones que hacen que el sistema sea discontinuo. Este método es el recomendado por MATLAB para resolver este problema [The MathWorks [n.d.]].

3.1 Condiciones iniciales

Para que ode45 pueda resolver el sistema de ecuaciones diferenciales hay que proporcionarle un vector de condiciones iniciales, es decir, los individuos en cada compartimento al inicio del periodo de resolución.

En este modelo, suponemos que en todos los compartimentos hay 0 personas excepto en:

- **Susceptibles:** Todos los individuos de la población a modelar menos los latentes y los infectados
- **Latentes:** Un parámetro del modelo a optimizar ($L(0)$).
- **Infectados:** Supondremos que en todas las comunidades habrá dos infectados al inicio del periodo de resolución, a falta de un modelo más robusto para modelar las interacciones entre comunidades.

4 VALORES DEL MODELO CON PARÁMETROS FICTICIOS

Para comprobar que el modelo planteado se comporta según lo esperado, este se ha resuelto numéricamente con los parámetros de la Tabla 1.

Como se puede observar en las Figuras 4, 5, 6, 7 y 8 el modelo se comporta según lo esperado y concluimos que el modelo, con unos parámetros correctamente optimizados, podría predecir correctamente la evolución de la enfermedad.

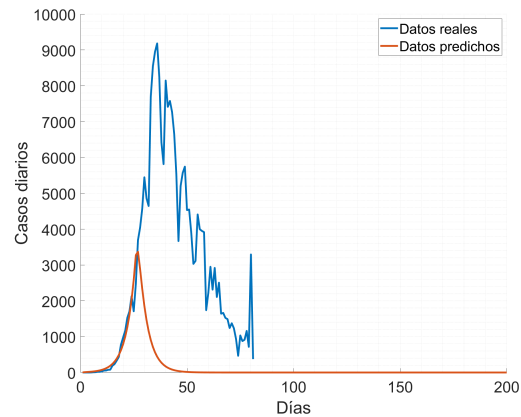


Fig. 4. Gráfico del comportamiento del modelo propuesto con los parámetros ficticios respecto a los casos diarios.

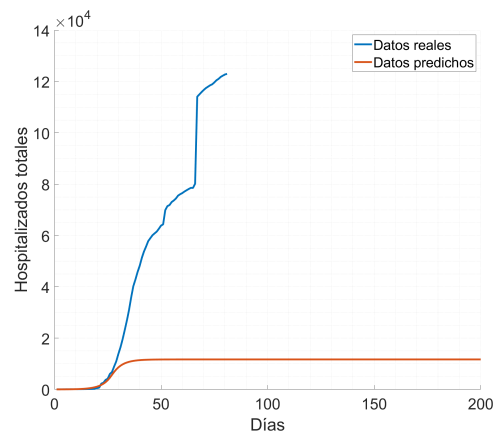


Fig. 5. Gráfico del comportamiento del modelo propuesto con los parámetros ficticios respecto a los hospitalizados totales.

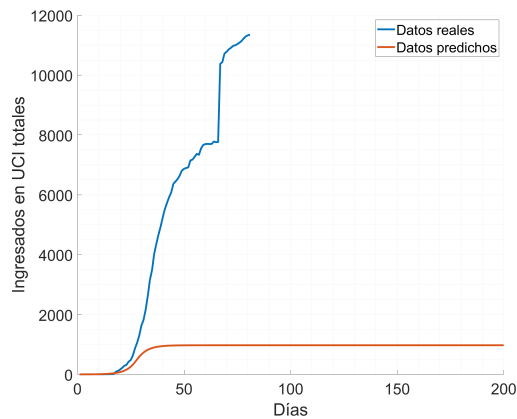


Fig. 6. Gráfico del comportamiento del modelo propuesto con los parámetros ficticios respecto a ingresados en UCI totales.

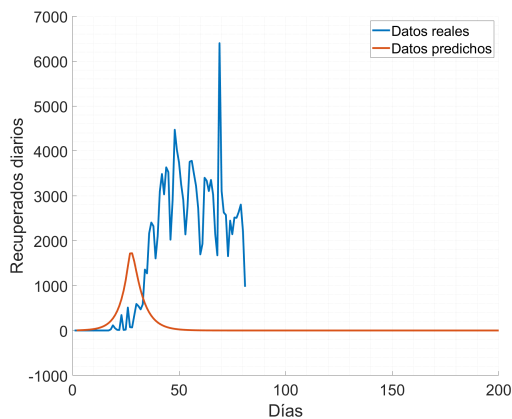


Fig. 7. Gráfico del comportamiento del modelo propuesto con los parámetros ficticios respecto a recuperados diarios.

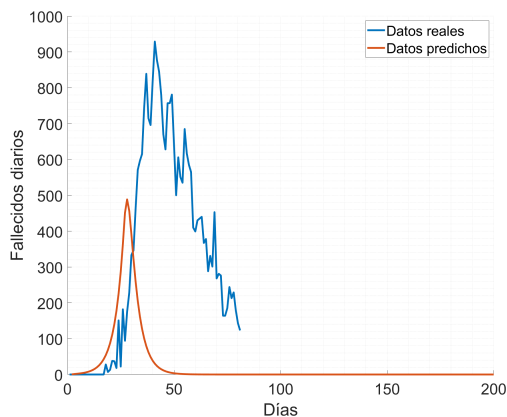


Fig. 8. Gráfico del comportamiento del modelo propuesto con los parámetros ficticios respecto a fallecidos diarios.

REFERENCES

- Ottar N. Bjørnstad, Bärbel F. Finkenstädt, and Bryan T. Grenfell. 2002. Dynamics of measles epidemics: Estimating scaling of transmission rates using a Time series SIR model. *Ecological Monographs* 72, 2 (2002), 169–184. [https://doi.org/10.1890/0012-9615\(2002\)072\[0169:DOMEES\]2.0.CO;2](https://doi.org/10.1890/0012-9615(2002)072[0169:DOMEES]2.0.CO;2)
- Herbert W. Hethcote. 2008. The basic epidemiology models: models, expressions for R_0 , parameter estimation, and applications. (12 2008). https://doi.org/10.1142/9789812834836_0001
- William Ogilvy Kermack and A. G. McKendrick. 1927. A contribution to the mathematical theory of epidemics. 115 (08 1927). <https://doi.org/10.1098/rspa.1927.0118>
- Leonardo R Lopez and Xavier Rodo. 2020. A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics. *medRxiv* (2020). <https://doi.org/10.1101/2020.03.27.20045005> arXiv:<https://www.medrxiv.org/content/early/2020/04/16/2020.03.27.20045005.full.pdf>
- Organización Mundial de la Salud. 2020. "Immunity passports in the context of COVID-19". Retrieved 10 Mayo 2020 from <https://www.who.int/news-room/commentaries/detail/immunity-passports-in-the-context-of-covid-19>
- The MathWorks. [n.d.]. "Troubleshoot Common ODE Problems". Retrieved 11 Mayo 2020 from <https://www.mathworks.com/help/matlab/math/troubleshoot-common-ode-problems.html>

AGRADECIMIENTOS

Queremos agradecer a Daniel Nieto, alumno de Máster de Ingeniería Industrial por la ayuda prestada con las dudas que hemos tenido a la hora de enfrentarnos con sistemas de ecuaciones diferenciales y su resolución numérica.