

Ajuste de parámetros de un modelo SIR modificado para la 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. Ajuste de parámetros de un modelo SIR modificado para la predicción de la COVID-19 en España, Desarrollo de Sistemas Inteligentes. 1, 1 (June 2020), 7 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.

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.

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:

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

2 MODELO EPIDEMIOLÓGICO PLANTEADO

En el modelo epidemiológico inicialmente planteado intentamos incorporar todas las mejoras anteriormente propuestas al modelo SIR básico mientras que añadíamos otros compartimentos específicos del comportamiento de la COVID-19 en España, como por ejemplo los infectados asintomáticos, hospitalizados ingresados en UCI, recuperados que estuvieron ingresados en UCI o recuperados asintomáticos

En el diseño de este modelo no se tuvieron en cuenta las dinámicas vitales de la población, ya que la evolución de la epidemia de la COVID-19 en España alcanzó el pico de contagios en menos de dos meses, por lo que se estimó 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. Los compartimentos están detallados en [Cordoba and Corroto 2020].

Mientras intentábamos optimizar los parámetros del modelo nos dimos cuenta de varios errores que habíamos cometido a la hora de diseñar el modelo, antes de intentar adaptar el modelo a los datos disponibles en España detallaremos los errores encontrados y corregidos en este nuevo modelo modificado:

- **Los individuos recuperados asintomáticos no cuentan en las estadísticas de recuperados**, ya que no han desarrollado síntomas ni han precisado de hospitalización, por lo que estos individuos deberán ir a otro compartimento, denominado **RA**.
- **Los individuos asintomáticos también son capaces de infectar**

El diagrama de estados de este modelo corregido se puede ver en la Figura 1.

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/2020/6-ART \$15.00

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

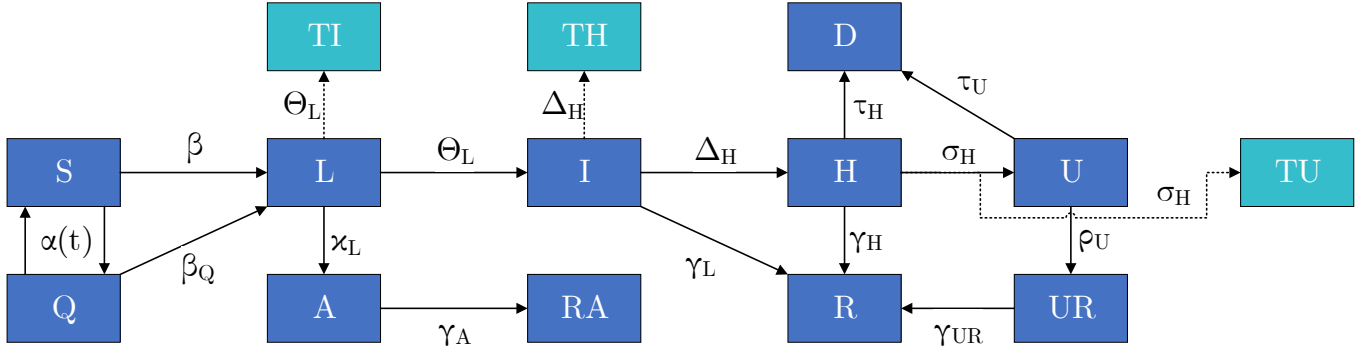


Fig. 1. Diagrama de estados del modelo epidemiológico propuesto modificado. Los compartimentos en color cian son los compartimentos añadidos para poder adaptar el modelo a los datos proporcionados por el Ministerio de Sanidad.

2.1 Modificaciones para adaptar el modelo a los datos de España

Los datos que proporciona el primer modelo que propusimos en [Cordoba and Corroto 2020] no son compatibles con los datos proporcionados por el Ministerio de Sanidad, por ello, y para poder obtener datos más interpretables tuvimos que modificarlo ligeramente.

El problema que surge con el modelo, es que al resolver el sistema de ecuaciones se obtiene una solución que representa el número de individuos que había en cada compartimento, sin embargo, los datos proporcionados por el Ministerio son acumulados, por lo que no se podrían comparar ambos datos y, como consecuencia, tampoco se podría calcular una función de pérdida para el método de optimización.

Como se comentó en [Cordoba and Corroto 2020], en estos modelos hay dos tipos de compartimentos, intermedios y finales. Los compartimentos finales se comportan como un acumulado respecto al tiempo, por lo que si consiguiésemos convertir estos compartimentos intermedios (p.e. Infectados u Hospitalizados) en compartimentos finales habríamos resuelto el problema.

Con esta finalidad, se añaden tres nuevos compartimentos finales, TI, TH y TU, las tasas de entrada de individuos a estos compartimentos son las mismas que sus homólogos intermedios, solo que sin tasas de salida.

En la Figura 15 se puede observar la comparación entre los valores en el tiempo del compartimento intermedio (Infectados) y en el final (TI).

El sistema de ecuaciones diferenciales del modelo modificado está descrito debajo y consta de 14 compartimentos y 16 parámetros.

$$\begin{aligned}\frac{dQ}{dt} &= \alpha(t) * S - \frac{\beta_Q * Q * (I + A)}{N_c} \\ \frac{dS}{dt} &= -\alpha(t) * S - \frac{\beta * S * (I + A)}{N_c} \\ \frac{dL}{dt} &= \frac{\beta_Q * Q * (I + A)}{N_c} + \frac{\beta * S * (I + A)}{N_c} - \theta_L * L - \kappa_L * L\end{aligned}$$

$$\begin{aligned}\frac{dI}{dt} &= \theta_L * L - \delta_I * I - \gamma_I * I \\ \frac{dA}{dt} &= \kappa_L * L - \gamma_A * A \\ \frac{dH}{dt} &= \delta_I * I - \gamma_H * H - \tau_H * H - \sigma_H * H \\ \frac{dU}{dt} &= \sigma_H * H - \tau_U * U - \rho_U * U \\ \frac{dUR}{dt} &= \rho_U * U - \gamma_{UR} * UR \\ \frac{dR}{dt} &= \gamma_I * I + \gamma_H * H + \gamma_{UR} * UR \\ \frac{dD}{dt} &= \tau_H * H + \tau_U * U \\ \frac{dRA}{dt} &= \gamma_A * A \\ \frac{dTI}{dt} &= \theta_L * L \\ \frac{dTH}{dt} &= \delta_I * I \\ \frac{dTU}{dt} &= \sigma_H * H\end{aligned}$$

3 OPTIMIZACIÓN DE LOS PARÁMETROS DEL MODELO

Una vez el modelo ha sido diseñado habrá que encontrar los valores correctos de los parámetros para que los valores predichos por el modelo sean los más cercanos posibles a los valores reales. Estos valores reales corresponden a las siguientes variables:

- Casos diarios
- Hospitalizaciones acumuladas
- Ingresos en UCI acumulados
- Recuperaciones acumuladas
- Defunciones acumuladas

La función de pérdida a optimizar será la Raíz del Error Cuadrático Medio (RECM / RMSE), probando también la Raíz del Error Cuadrático Medio Normalizado (RECMN / NRMSE)

Se han planteado dos enfoques diferentes para optimizar el modelo:

- El uso de un optimizador iterativo no lineal con restricciones con un único valor de pérdida con un único objetivo.
- El uso de un optimizador multiobjetivo con restricciones para encontrar la frontera de Pareto, aquella en la cual se encuentra el conjunto de todas las soluciones óptimas en el sentido de Pareto.

3.1 Optimización uniobjetivo

La optimización uniobjetivo consiste en encontrar el valor de la función que hace que el valor de esta sea lo más pequeño (o más grande) posible.

El principal inconveniente de este enfoque es que únicamente se puede tener en cuenta un objetivo para optimizar la función. Sin embargo, en nuestro modelo contamos con 5 objetivos, uno por cada dato disponible.

Para resolver este problema, hemos sumado todas las RECM en una única RECM global para así proporcionar el único objetivo que espera el optimizador.

3.1.1 Problema de los mínimos locales. Los optimizadores uniobjetivo más utilizados emplean el método del descenso del gradiente, un método propenso a estancarse en mínimos locales, sin embargo, en nuestra formulación desearíamos obtener el mínimo global, es decir, el mejor ajuste posible del modelo a los datos disponibles.

El método del descenso del gradiente sólo se puede aplicar a funciones diferenciables (por esta razón hemos escogido una función como RECM como función de pérdida). El método consiste en ir tomando iterativamente pasos en la dirección del gradiente del punto actual, hasta que el punto no cambie entre iteraciones o el gradiente sea cero.

Sin embargo, no es condición suficiente que el gradiente sea cero o que el punto no cambie entre iteraciones para asegurar que el método haya llegado al mínimo global.

3.1.2 Métodos de optimización global. Hay muchos enfoques a la hora de resolver el problema de los mínimos locales, el que hemos empleado en este trabajo ha sido un algoritmo genético y el método de enjambre de partículas.

En ambos casos, los métodos intentan encontrar el mínimo global creando nuevas partículas en posiciones aleatorias del espacio de búsqueda y modificando sus posiciones dependiendo del rendimiento del resto de partículas (enjambre de partículas) o creando nuevas partículas a partir de las más adaptadas (algoritmo genético).

Estos métodos tienen dos problemas, el primero de ellos es que no son deterministas debido a su inicialización aleatoria. Este problema se puede resolver fácilmente estableciendo siempre la misma semilla del generador de números aleatorios al mismo valor.

El segundo problema es su precisión reducida, al generar nuevas partículas de forma aleatoria su velocidad de convergencia es mucho menor comparada con los métodos de descenso del gradiente.

3.1.3 Métodos de optimización híbridos. Sin embargo, hay una forma de combinar ambos métodos para obtener lo mejor de ambos, la habilidad de llegar a mínimos globales y la precisión aumentada.

Estos métodos consisten en ejecutar un método de descenso del gradiente una vez un método de optimización global ha concluido su ejecución.

Este el método escogido para realizar la optimización uniobjetivo en el modelo.

3.2 Optimización multiobjetivo

A pesar de poder formular el problema de optimización como uno uniobjetivo el modelo se presta desde el inicio a una formulación multiobjetivo, donde cada variable es considerada un objetivo.

Como se ha explicado anteriormente, estos métodos intentan encontrar la frontera de Pareto, es decir, aquella en la que se encuentra todo el conjunto de soluciones óptimas en el sentido de Pareto.

Este método también se presta a ser parte de un método híbrido como se ha comentado en la sección anterior.

Una vez revisados los métodos de optimización, hemos decidido que probaremos las siguientes combinaciones de métodos y funciones de pérdida:

- (1) Optimización uniobjetivo
 - (a) Con función de pérdida RECM
 - (b) Con función de pérdida RECMN
- (2) Optimización multiobjetivo
 - (a) Con función de pérdida RECM
 - (b) Con función de pérdida RECMN

A los parámetros se les impuso límites inferiores de 0 y superiores de 1 para el porcentaje de individuos que transicionan a cuarentena, infinito para los latentes iniciales y 2 para el resto de parámetros.

A los métodos de optimización no se les impondrá ninguna restricción temporal o de evaluación de funciones o iteraciones, con el objetivo de que sean capaces de llegar a la solución más óptima posible.

Ya que no impondremos restricciones temporales será necesario paralelizar la ejecución de los métodos para que los tiempos de ejecución no se alarguen demasiado. Como ejemplo, una ejecución sin paralelizar tardó en ejecutarse más de 20 minutos.

4 RESULTADOS

Los resultados de la optimización multiobjetivo, tanto con la RECM como con la RECMN son deplorables, no consiguen acercarse a los datos reales, por lo que desechamos la idea de escoger la optimización multiobjetivo como método final. Esto se puede ver en las Figuras 9, 10, 11, 12 y 13

Por lo que nos quedaría comparar la optimización uniobjetivo con la RECM y la RECMN.

Hemos escogido la RECM como métrica para comparar ambas optimizaciones: La optimización uniobjetivo con la RECM obtuvo **19265.28** mientras que la optimización con la RECMN obtuvo **30877.84** por lo podemos concluir que el modelo optimizado con un método uniobjetivo con función de pérdida RECM es la que más se acerca a los datos proporcionados.

Por último, el uso de la RECM como métrica para comparar los modelos podría no ser la óptima, ya que el modelo optimizado con la RECM priorizaría reducir el error en los datos con mayor rango. Sin embargo, consideramos que este supuesto no se aplica ya que la diferencia entre los valores de las métricas es lo suficientemente grande para considerar este efecto como la causa de la misma.

5 CONOCIMIENTO EXTRAÍDO DEL MODELO

Una vez optimizados los parámetros, y conociendo los parámetros más importantes, en la Tabla 1, podemos modificar uno o varios parámetros para conocer como sería la evolución de la enfermedad en otras condiciones.

Table 1. Tabla con los parámetros más importantes del modelo ya optimizado.

Parámetro	$\alpha(t)$	$L(0)$	β_B	β_A	β_Q
Valor	0.3417	18.2	1.43	1.48	0.0004
Parámetro	γ_A	γ_I	γ_H	τ_H	τ_U
Valor	0.84	0.01	0.37	0.1558	0

Algunos ejemplos de modificaciones que podemos hacer son:

- Atrasar o adelantar el día en el que la cuarentena se aplicó para comprobar cómo de efectiva ha sido la misma.
- Modificar el número de infectados latentes para comprobar cómo se comportaría la enfermedad si se hubiesen trazado mejor los contactos con los infectados.
- Añadir desescaladas coincidiendo con los días de las fases para predecir a priori cómo se comportará la enfermedad.

Para este trabajo hemos escogido la primera opción, simularemos cómo se habría comportado la enfermedad si se hubiese declarado la cuarentena 2 semanas antes, el día 1 de marzo.

Los resultados de esta simulación se pueden ver en las Figuras 2, 3, 4, 5 y 6.

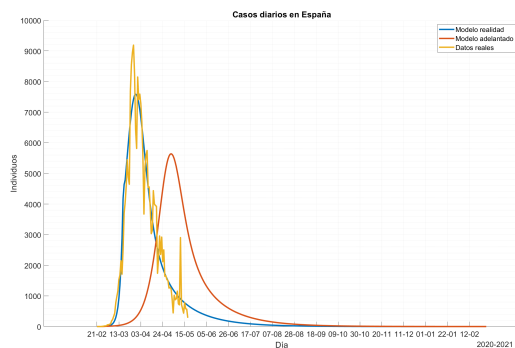


Fig. 2. Gráfico del modelo optimizado usando la RECM y optimizador uniobjetivo. Los datos corresponden a los casos diarios reportados.

Otra hipótesis que hemos testado es el grado de involucración de los infectados asintomáticos en la evolución de la enfermedad, para ello, hemos eliminado la capacidad de los asintomáticos de infectar manteniendo el resto del modelo sin modificaciones.

A la hora de intentar optimizar el modelo, el RECM ha ascendido hasta **34178.83**, más incluso que en el caso de optimizarlo con RECMN. Como se puede ver en la Figura 7, en la gráfica de los casos diarios, el modelo no consigue acercarse a los datos reales proporcionados, e incluso en el caso de adelantar la cuarentena el modelo deja de funcionar.

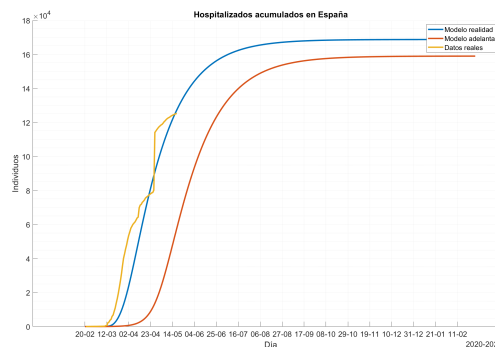


Fig. 3. Gráfico del modelo optimizado usando la RECM y optimizador uniobjetivo. Los datos corresponden a los hospitalizados acumulados reportados.

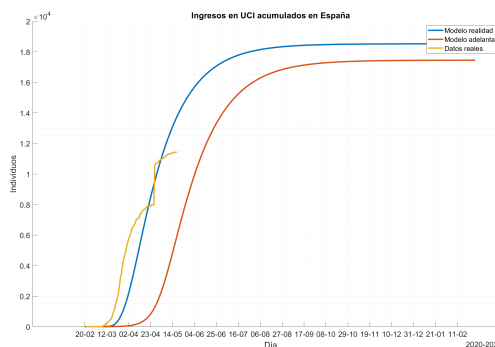


Fig. 4. Gráfico del modelo optimizado usando la RECM y optimizador uniobjetivo. Los datos corresponden a los ingresos en UCI acumulados reportados.

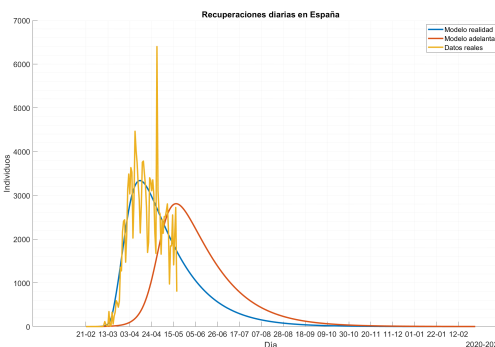


Fig. 5. Gráfico del modelo optimizado usando la RECM y optimizador uniobjetivo. Los datos corresponden a las recuperaciones diarias reportadas.

5.1 Conclusiones

Como se puede observar en las figuras mencionadas anteriormente, el modelo consigue predecir correctamente las 5 variables, aunque

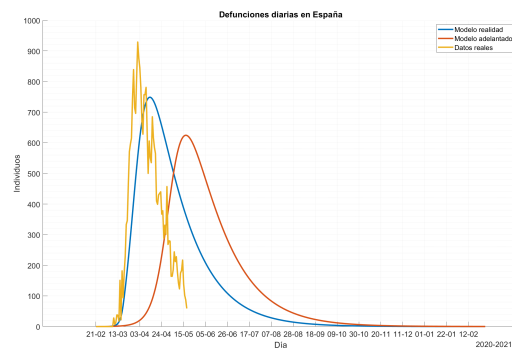


Fig. 6. Gráfico del modelo optimizado usando la RECM y optimizador uniobjetivo. Los datos corresponden a las defunciones diarias reportadas.

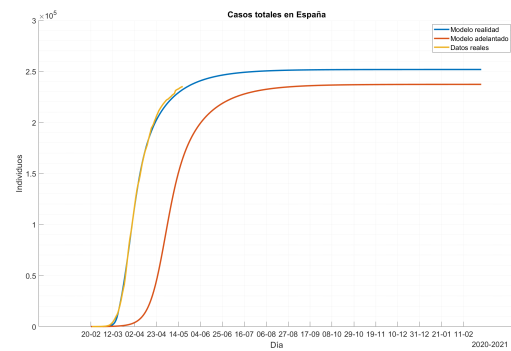


Fig. 8. Gráfico del modelo optimizado usando la RECM y optimizador uniobjetivo. Los datos corresponden a los casos totales reportados.

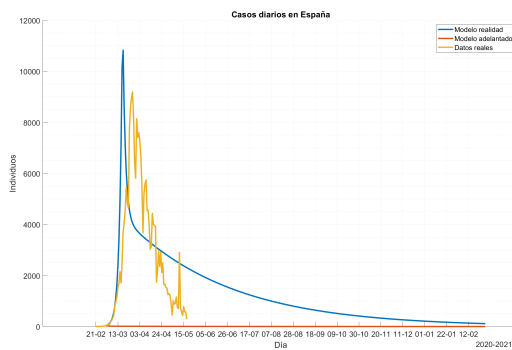


Fig. 7. Gráfico del modelo optimizado eliminando a los asintomáticos la capacidad de infectar. Los datos corresponden a los casos diarios reportados.

A GRÁFICAS COMPLEMENTARIAS

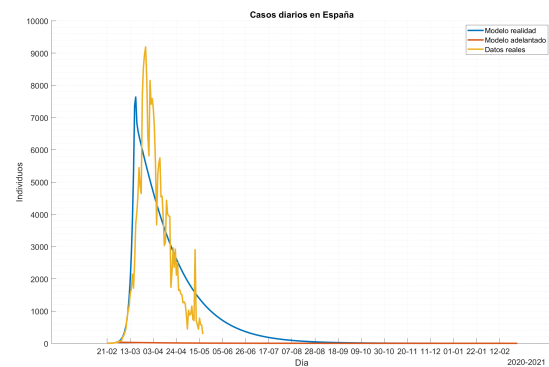


Fig. 9. Gráfico del modelo optimizado usando la RECMN y optimizador uniobjetivo. Los datos corresponden a los casos diarios reportados.

en el caso de las hospitalizaciones e ingresos en UCI el modelo sobreestima ligeramente el crecimiento de estas variables.

Aunque, si nos ponemos en la situación actual de la enfermedad en España, esta sobreestimación podría deberse al descuadre de datos que motivó el ajuste de las series temporales hace una semana [EFE 2020], o a las supuestas defunciones incorrectamente notificadas.

Otra conclusión que podemos extraer, es que en el caso de haberse declarado el confinamiento dos semanas antes, actualmente estaríamos en el mismo estado en el que estábamos en el 28 de Abril, con 1700 nuevos casos diarios. Aunque como se puede ver en la Figura 8 los casos totales bajarían de 250 000 a 237 000.

La conclusión que se extrae de estos datos es que la acción temprana es necesaria para parar de forma eficaz una pandemia como ha sido la del COVID-19.

REFERENCES

- Javier Cordoba and Juan Jose Corroto. 2020. Modelo epidemiológico de predicción de la COVID-19 en España. (2020).
- EFE. 2020. "Los cambios de criterio en el recuento de víctimas del coronavirus". Retrieved 01 Junio 2020 from <https://www.efc.com/efe/espana/sociedad/los-cambios-de-criterio-en-el-recuento-victimas-del-coronavirus/10004-4255595>
- 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>

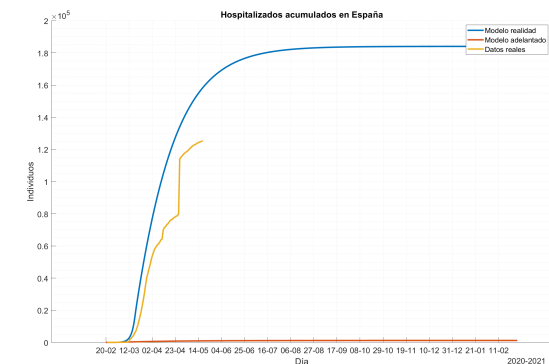


Fig. 10. Gráfico del modelo optimizado usando la RECMN y optimizador uniobjetivo. Los datos corresponden a los hospitalizados acumulados reportados.

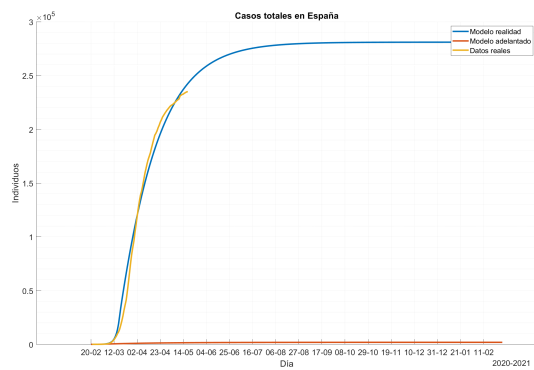


Fig. 14. Gráfico del modelo optimizado usando la RECMN y optimizador uniobjetivo. Los datos corresponden a los casos totales reportados.

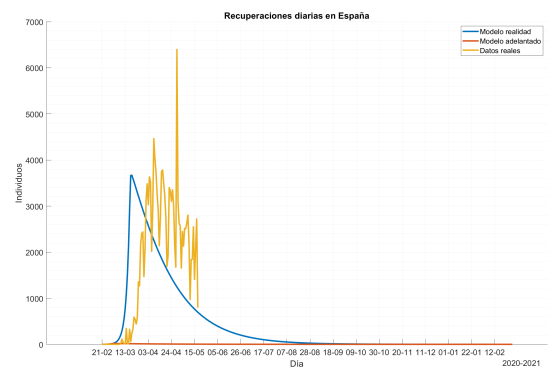


Fig. 12. Gráfico del modelo optimizado usando la RECMN y optimizador uniobjetivo. Los datos corresponden a las recuperaciones diarias reportadas.

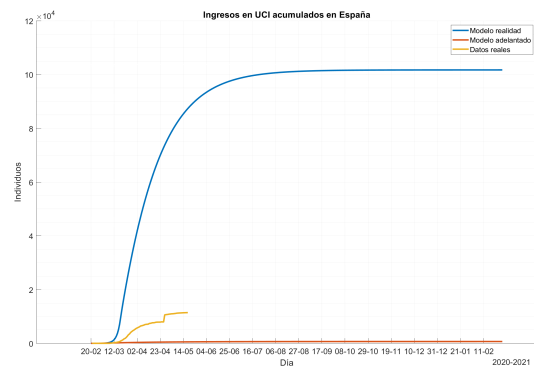


Fig. 11. Gráfico del modelo optimizado usando la RECMN y optimizador uniobjetivo. Los datos corresponden a los ingresos en UCI acumulados reportados.

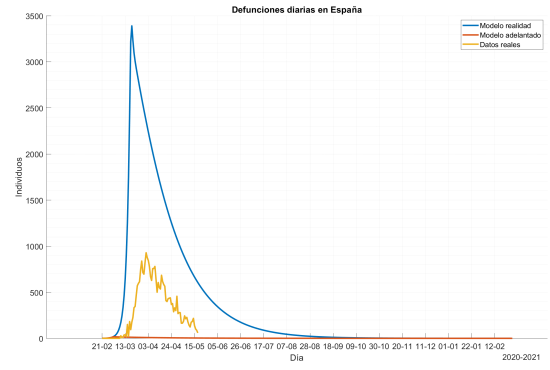


Fig. 13. Gráfico del modelo optimizado usando la RECMN y optimizador uniobjetivo. Los datos corresponden a las defunciones diarias reportadas.

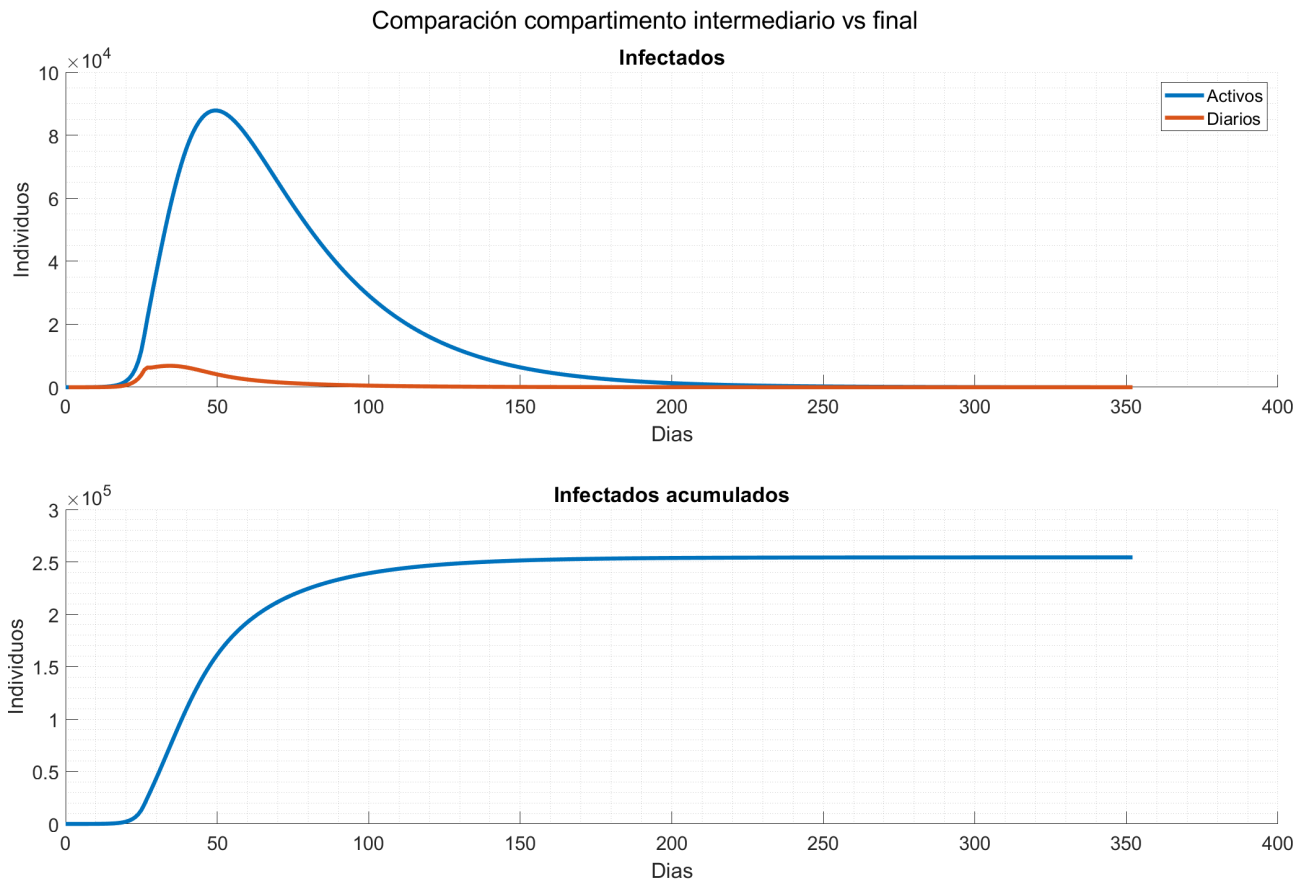


Fig. 15. Gráfico del comportamiento del compartimento intermedio frente al de un compartimento final (Infectados).