



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Modelo matemático-computacional de la evolución de un
cáncer de mama en una paciente tras su resección

Trabajo Fin de Grado

Grado en Ingeniería Informática

AUTOR/A: Villanueva Tarazona, José

Tutor/a: Villanueva Micó, Rafael Jacinto

Cotutor/a: Hervás Jorge, Antonio

CURSO ACADÉMICO: 2023/2024

Resumen

Este trabajo de fin de grado surgió de una colaboración con un equipo del Instituto Universitario de Matemática Multidisciplinar de la Universitat Politècnica de València y un equipo médico formado por personal del Hospital Clínico Universitario y del Hospital Universitario Doctor Peset. Dicha colaboración tiene como objetivo cuantificar el retraso en la aparición de recurrencias en cáncer de mama en mujeres en la que se administra quimioterapia preventiva tras extirparles el tumor. Este objetivo lo vamos a realizar mediante la construcción de un modelo en dos etapas. La etapa previa a la resección del tumor de mama, en la que la paciente tiene un tumor que se trata administrando quimioterapia para reducir su tamaño antes de proceder a la resección. La segunda etapa, tras la resección, consistirá en modelizar el crecimiento tumoral hasta el punto en el que se considere que ocurre una recurrencia. Una vez determinados los parámetros de dicho modelo se propone realizar simulaciones del modelo y determinar el momento óptimo para administrar sesiones de quimioterapia para retrasar la recurrencia lo máximo posible.

Concretamente, el TFG tiene como objetivo el desarrollo de un prototipo que nos permita implementar el modelo construido, calibrarlo a los datos de un único paciente, hacer una proyección de cuándo aparecerá la recurrencia, simular posibles ciclos de quimioterapia después de la resección y maximizar el retraso de la recurrencia.

Resum

Aquest treball de fi de grau va sorgir d'una col·laboració amb un equip matemàtic de l'Institut Universitari de Matemàtica Multidisciplinària de la Universitat Politècnica de València i un equip mèdic format per personal de l'Hospital Clínic Universitari i de l'Hospital Universitari Doctor Peset. Aquesta col·laboració té com a objectiu quantificar el retard en l'aparició de recurrències en càncer de mama en dones en la qual s'administra quimioteràpia preventiva després d'extirpar-los el tumor. Este objectiu el realitzarem mitjançant la construcció d'un model en dues etapes. L'etapa prèvia a la resecció del tumor de mama, en la qual la pacient té un tumor que es tracta administrant quimioteràpia per a reduir la seua grandària abans de procedir a la resecció. La segona etapa, després de la resecció, consistirà a modelitzar el creixement tumoral fins al punt en el qual es considere que ocorre una recurrència. Una vegada determinats els paràmetres d'aquest model es proposa realitzar simulacions del model i determinar el moment òptim per a administrar sessions de quimioteràpia per a retardar la recidiva el màxim possible.

Concretament, el TFG té com a objectiu el desenvolupament d'un prototip que ens permeta implementar el model construït, calibrar-lo a les dades d'un únic pacient, fer una projecció de quan apareixerà la recurrència, simular possibles cicles de quimioteràpia després de la resecció i maximitzar el retard de la recurrència.

Abstract

This final degree project arose from a collaboration with a mathematical team from the *Instituto Universitario de Matemática Multidisciplinar de la Universitat Politècnica de València* and a medical team made up of staff from the *Hospital Clínico Universitario* and the *Hospital Universitario Doctor Peset*. The aim of this collaboration is to quantify the delay in the appearance of recurrences in breast cancer in women in whom preventive chemotherapy is administered after removal of the tumor. This objective will be achieved through the construction of a two-stage model. The stage prior to resection of the breast tumour, in which the patient has a tumour that is treated by administering chemotherapy to reduce its size before proceeding to resection. The second stage, after resection, will consist of modelling tumour growth to the point at which recurrence is considered to occur. Once the parameters of the model have been determined, it is proposed to run simulations of the model and determine the optimal time to administer chemotherapy sessions to delay recurrence as long as possible.

Specifically, the TFG aims to develop a prototype that allows us to implement the model built, calibrate it to data from a single patient, project when recurrence will occur, simulate possible cycles of chemotherapy after resection and maximise the delay of recurrence.

Palabras clave

Modelado matemático-computacional, Cáncer de mama, HER2, Resección, Recurrencia, Recidiva, Evolución, Incertidumbre, Particle Swarm Optimization.

Índice

Resumen.....	1
Resum.....	2
Abstract.....	3
Palabras clave.....	3
Capítulo 1. Introducción.....	6
1.1. Aproximación al problema.....	6
El cáncer.....	6
Diferencias entre las células cancerosas y las células normales.....	7
Cáncer de mama.....	7
Cáncer de mama HER2+.....	10
1.2. Colaboraciones.....	11
1.3. Metodología.....	11
1.4. Objetivos.....	12
1.5. Estructura.....	13
Capítulo 2. Estado del arte.....	14
2.1. Revisión bibliográfica.....	14
2.2. Propuesta.....	16
Capítulo 3. Análisis del problema.....	18
3.1. Datos disponibles.....	18
3.2. Herramientas disponibles.....	18
Modelización matemática.....	21
3.3. Análisis del marco legal y ético.....	22
3.4. Solución propuesta.....	23
Capítulo 4. Desarrollo de la solución propuesta.....	26

4.1. El primer script.....	29
4.2. El segundo script.....	30
4.3. El tercer script.....	32
Capítulo 5. Pruebas.....	34
Capítulo 6. Conclusiones.....	38
6.1.Relación del trabajo desarrollado con los estudios cursados.....	39
Anexo ODS.....	43
OBJETIVOS DE DESARROLLO SOSTENIBLE.....	43

Capítulo 1. Introducción

1.1. Aproximación al problema

El cáncer

El cáncer es una enfermedad por la que varias células del cuerpo se multiplican sin control y se propagan a otras partes del cuerpo y que, de no tratarse, pueden causar la muerte.¹

El cáncer puede comenzar en cualquier parte del cuerpo. Normalmente las células humanas se forman y se multiplican (mediante un proceso llamado división celular) para formar células nuevas según las va necesitando el cuerpo. Cuando las células se dañan, mueren y las células nuevas las reemplazan.

A veces el proceso no sigue este orden y las células dañadas se forman y se multiplican cuando no deberían. Estas células pueden formar tumores. Los tumores pueden ser cancerosos (malignos) o no cancerosos (benignos):

- Los tumores cancerosos, cuando son sólidos, son capaces de invadir los tejidos cercanos. También puede ocurrir una metástasis, que consiste en que el las células viajan a otras partes del cuerpo y forman otros tumores. La mayoría de los tipos de cáncer son sólidos, con la excepción de cánceres como la leucemia, que no forman tumores sólidos.
- Esto no ocurre con los tumores benignos. Cuando se extirpan los tumores benignos, no suelen volver, mientras que los tumores cancerosos a veces vuelven. Los tumores benignos pueden llegar a alcanzar tamaños bastante grandes, y podrían llegar a causar síntomas graves o poner en peligro la vida de la persona, si apareciesen en lugares delicados, como en el cerebro o el encéfalo.

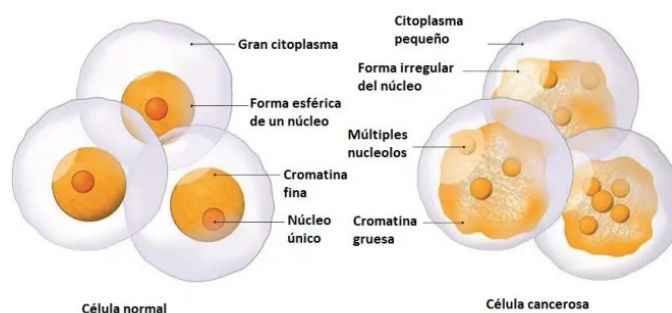


Figura 1: Diferencia entre célula normal y célula cancerosa. Fuente: <https://i1.wp.com/www.elgencurioso.com/wp-content/uploads/2021/05/13-celula-normal-contra-cancerosa.jpg?resize=768%2C329&ssl=1>

1 <https://www.cancer.gov/espanol/cancer/naturaleza/que-es>

Diferencias entre las células cancerosas y las células normales

Las células cancerosas difieren de las células normales en que²:

- Las células normales solo se forman cuando reciben señales de que se deben formar.
- Las células cancerosas no hacen caso a las señales que indican a las células que dejen de multiplicarse o que deben destruirse.
- Las células normales dejan de multiplicarse cuando se encuentran con otras células, y las células normales no suelen moverse a otras partes del cuerpo.
- Las células cancerosas hacen que los vasos sanguíneos crezcan en dirección a los tumores, que les llevan oxígeno y nutrientes.
- Las células cancerosas se esconden del sistema inmunitario. El sistema inmunitario suele destruir las células dañadas o anormales.
- Las células cancerosas engañan al sistema inmunitario para seguir vivas y multiplicarse. Algunas células cancerosas incluso logran que las células inmunitarias protejan el tumor en vez de atacarlo.
- Las células cancerosas incorporan numerosos cambios en los cromosomas, como partes del cromosoma que se repiten o faltan. Algunas células cancerosas tienen el doble de cromosomas.
- Las células cancerosas dependen de nutrientes que son distintos a los que necesitan las células normales. Además, algunas células cancerosas crean energía de los nutrientes de formas diferentes a las que usan la mayoría de las células normales. Esto permite que se multipliquen mucho más rápido.

Muchas veces, las células cancerosas necesitan muchísimo de estos comportamientos anormales, ya que no pueden sobrevivir sin estos. Esto permite a los investigadores idear terapias que tengan como objetivo a células que exhiban estos comportamientos. Por ejemplo, algunas terapias de cáncer evitan que los vasos sanguíneos crezcan en dirección a los tumores, y esto evita que el tumor consiga los nutrientes que necesita.

Cáncer de mama

En 2020 se diagnosticó cáncer de mama a 2,3 millones de mujeres en todo el mundo, de las cuales 685.000 fallecieron por su causa. A fines del mismo año, 7,8 millones de mujeres a las que en los cinco años anteriores se les había diagnosticado cáncer de mama seguían vivas, lo que convierte a este cáncer en el de mayor prevalencia en el mundo.³

2 <https://www.cancer.gov/espanol/cancer/naturaleza/que-es>

3 <https://www.who.int/es/news-room/fact-sheets/detail/breast-cancer>

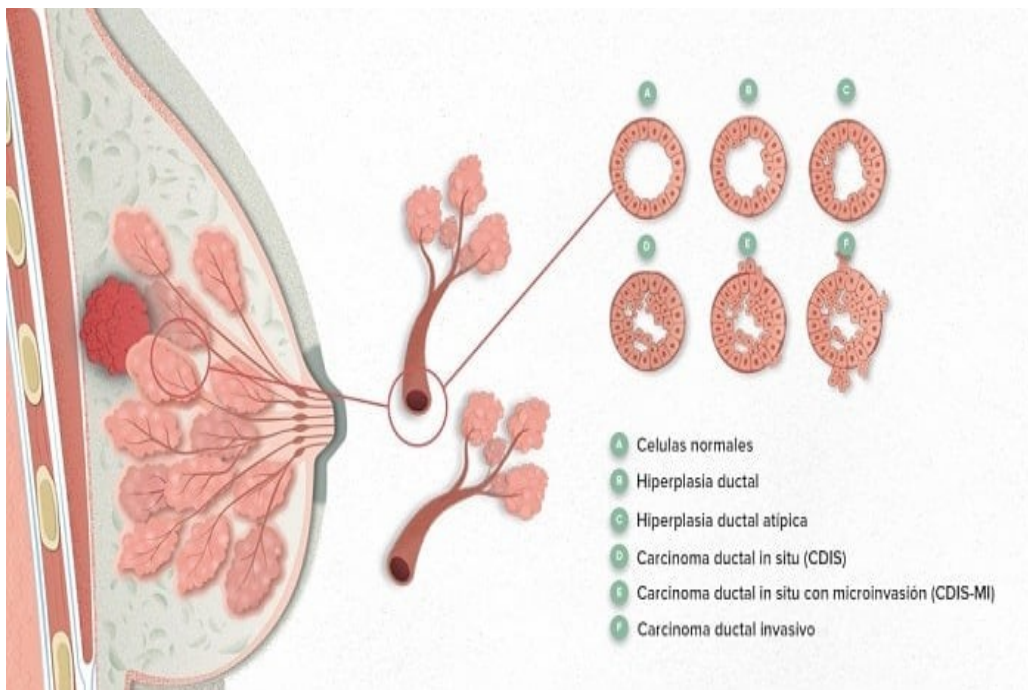


Figura 2: Diagrama de un cáncer de mama. Los distintos números indican distintos tipos de cáncer. Fuente: <https://encolombia.com/wp-content/uploads/2015/02/carcinoma-ductal-mama.jpeg>

Evolución de los casos de cáncer de mama cada año en España

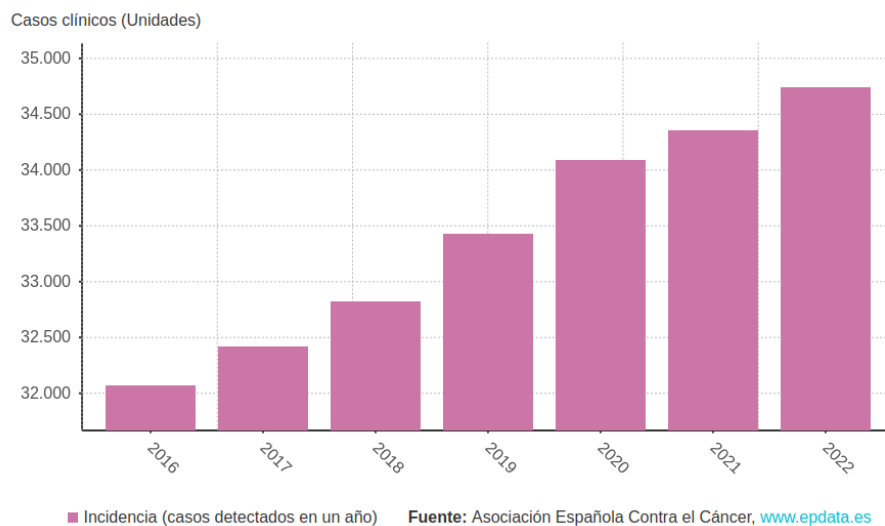


Figura 3: Gráfica que representa el número de casos clínicos de cancer de mama por año. Fuente: www.epdata.es

Existen factores de riesgo que puedan aumentar la posibilidad de padecer cáncer de mama, pero aproximadamente la mitad de todos los casos afectan a mujeres que no tienen factores de riesgo aparte del sexo y la edad. Aproximadamente entre el 0,5% y el 1% de los casos de cáncer de mama afectan a varones.

El cáncer de mama es una enfermedad en la cual las células de la mama se multiplican sin control. Existen distintos tipos de cáncer de mama, cuyo tipo depende de qué células de la mama se vuelven cancerosas.

Los tipos más comunes de cáncer de mama son:

- **Carcinoma ductal infiltrante.** Las células cancerosas se originan en los conductos y después salen de ellos y se multiplican en otros tejidos mamarios. Estas células cancerosas invasoras también pueden diseminarse, o formar metástasis, en otras partes del cuerpo.
- **Carcinoma lobulillar infiltrante.** Las células cancerosas se originan en los lóbulos de la mama y después se diseminan de los lóbulos a los tejidos mamarios cercanos. En el caso de la mama, los lóbulos contienen los grupos de acinos glandulares, que son los que se encargan de producir la leche. Estas células cancerosas invasoras también pueden diseminarse a otras partes del cuerpo.

Existen otros tipos de cáncer de mama menos comunes, tales como los que se muestran en la Tabla 1:

Tabla 1: Tabla que indica la prevalencia de distintos carcinomas invasivos malignos. Fuente: [1]

Carcinoma invasivo (malignos)	Carcinoma ductal infiltrante	70%
	Carcinoma lobulillar infiltrante	4,46%
	Carcinoma medular	0,9%
	Carcinoma mucinoso o coloide	2,7%
	Carcinoma papilar infiltrante	2,7%
	Carcinoma tubular	2,7%
	Otros	16,54%

También ayuda clasificarlos por sus receptores, ya que nos indica lo idóneos que puede ser el uso de tratamientos específicos. Todas las células tienen receptores en su superficie. Ciertas hormonas se unen a dichos receptores, lo que provoca cambios en la célula, lo que les permite ser utilizados como marcadores tumorales. Los más comunes son los receptores de estrógeno, de progesterona, y del oncogén HER2. Nosotros nos vamos a centrar en el cáncer de mama positivo para HER2.

Cuando hablamos de cáncer, una **resección o tumorectomía** es un proceso con el objetivo de extirpar el cáncer u otro tejido anormal y una pequeña cantidad del tejido sano que lo rodea. Esto asegura que se elimine todo el tejido anormal. A diferencia de una mastectomía, en este proceso sólo se extrae una parte de la mama, en vez de todo el tejido mamario. La tumorectomía es una opción de tratamiento para personas con cáncer mamario en etapa temprana, y a veces se usa para descartar un diagnóstico de cáncer. Generalmente se realiza radioterapia en la mama después de la extracción del tumor para reducir las posibilidades de que vuelva a aparecer el cáncer.

Cuando se encuentra un cáncer después de un tratamiento, y después de un periodo de tiempo en el que el cáncer no se pudo detectar, se le llama una **recurrencia**. El cáncer puede volver a aparecer en el mismo sitio en el que apareció la primera vez, o aparecer en otro lugar completamente distinto. Aunque el tumor aparezca en otro sitio, sigue manteniendo el nombre de donde apareció por primera vez. Cuando el cáncer aparece en un lugar lejano de donde apareció, también se le llama metástasis.

Cáncer de mama HER2+

La HER2 es una proteína que es clave para el crecimiento y división normal de las células, por lo que su aparición en exceso está vinculada al cáncer de mama. Las células del cáncer de mama con niveles más altos de lo normal de HER2 se llaman *HER2-positivas*.

Aproximadamente el 25% de los cánceres de mama son de tipo HER2 positivo. Estos cánceres tienden a crecer y propagarse más rápido que otros tipos de cáncer de mama, es mucho más agresivo y tienen una mayor resistencia a los tratamientos convencionales. Sin embargo, responden al tratamiento con medicamentos que tienen como blanco a la proteína HER2, como el anticuerpo monoclonal, y combinando el trastuzumab con la quimioterapia convencional.^[2]

La **quimioterapia** es el tratamiento con medicamentos para interrumpir la formación de células cancerosas, ya sea mediante su destrucción o al impedir su multiplicación. La quimioterapia se administra por la boca, en inyección, por infusión o sobre la piel, según el tipo de cáncer y el estadio en que este se encuentra. Se administra sola o con otros tratamientos como cirugía, radioterapia o terapia biológica.

1.2. Colaboraciones

Nos integramos con un grupo de investigación del Instituto Universitario de Matemática Multidisciplinar con la idea de desarrollar un modelo matemático que sea capaz de predecir el tamaño de un cáncer de mama. Nos reunimos con un grupo de médicos del Hospital Clínico Universitario y Hospital Clínico Doctor Peset y nos propusieron la idea, que nos ayudaron a enfocar. Durante la reunión, se intercambiaron ideas sobre lo que el equipo matemático era capaz de modelar si se le proporcionaba información y datos, y la idea que se nos propuso fue la de usar el modelo para poder predecir cuándo aparece una recurrencia en un paciente.

El esquema de trabajo con el que funcionamos fue el de la Figura 4, donde el equipo matemático es el del Instituto Universitario de Matemática Multidisciplinar, el equipo médico corresponde al grupo formado por médicos del Hospital Clínico Universitario y el Hospital Clínico Doctor Peset, y un doctor del Servicio de Radiofísica del Hospital Clínico Universitario.

1.3. Metodología

Dividiremos el método de trabajo de la misma manera que lo hemos presentado en la Figura 4:

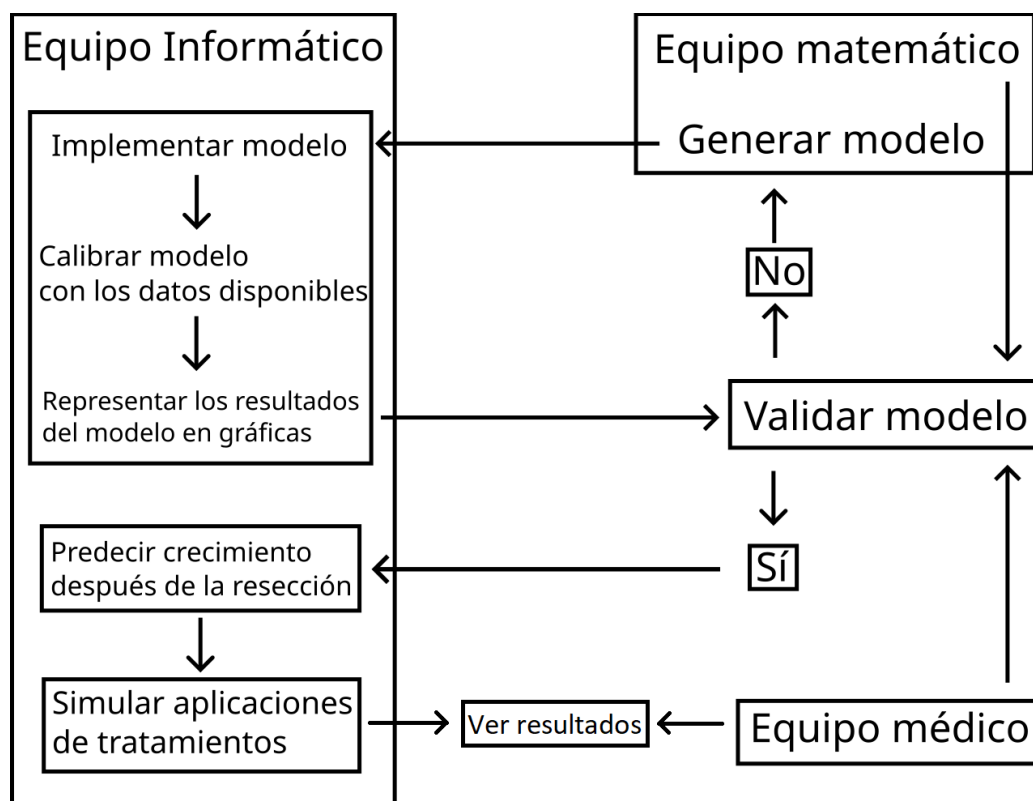


Figura 4: Esquema del flujo de trabajo. Los objetivos corresponden a las casillas individuales dentro del equipo informático

- **Implementar:** Dado el modelo generado por el equipo matemático, nosotros implementaremos el código del modelo matemático-computacional con técnicas de tratamiento de la incertidumbre en las que nos formará el equipo matemático. Una vez se pueda representar el modelo a lo largo de un paso temporal que especifiquemos, se pasará a la siguiente fase.
- **Calibrar:** Para calibrar el modelo usaremos técnicas de tratamiento de la incertidumbre, en las que nos formará el equipo matemático. Minimizaremos el error cuadrático medio entre los datos y el modelo. Esto debería devolvernos los valores de los parámetros calibrados, y con ellos ya podemos pasar a la siguiente fase y a la siguiente parte del problema.
- **Representar:** Una vez tengamos los parámetros del modelo, representaremos el modelo de crecimiento de cáncer de mama en una gráfica, que es la que mostraremos a los dos equipos para que validen los resultados.
- **Validar:** Por último, se le remite al equipo matemático los resultados de un informe gráfico para verificar que los resultados obtenidos se ajustan a la descripción del modelo generado. Si esto no fuera así, se modificará el modelo según indique el equipo matemático. El proceso se repite hasta obtener un modelo válido desde el punto de vista técnico. Entonces se realizará la validación con el equipo médico. Si el modelo es validado, entonces empezaremos a diseñar y ejecutar estrategias y simulaciones de predicción.

Una vez tengamos un modelo validado, entonces es cuando empezamos a hacer las predicciones con la ayuda del equipo matemático:

- **Predecir:** Haremos un predicción del crecimiento del tumor después de la resección, y observaremos cuánto tarda en ocurrir una recurrencia.
- **Simular:** Después, teniendo como base los datos obtenidos, los compararemos con los resultados que obtengamos de ir simulando los tratamientos que nos indique el equipo médico.

Una vez tengamos una serie de scripts que nos permita realizar fácilmente los bloques de calibrar, predecir y simular es cuando se podrá probar con otros pacientes.

1.4. Objetivos

El objetivo de este TFG es desarrollar un prototipo que nos permita cuantificar el retraso en la aparición de recurrencias en cáncer de mama en mujeres en la que se administra quimioterapia preventiva tras extirparles el tumor.

Este objetivo se puede separar en 3 subobjetivos, que son:

- Implementar un programa base que funcione usando el modelo generado por el equipo matemático, y podamos hacer una representación gráfica del mismo

antes de la resección, para poder aplicar un algoritmo de optimización para un paciente específico, y que nos devuelva los valores optimizados de dicho paciente.

- Implementar un programa base en el que poder usar los parámetros del modelo ya ajustados, indicarle el tamaño al que debería quedarse el tumor después de la resección para usarlo como tamaño inicial, y que nos devuelva el tamaño del tumor en cada instante sin quimioterapias aplicadas.
- Por último, que una vez tengamos el modelo después de la resección sin quimioterapias aplicadas, poder usarlo como base para comparar con el modelo con simulaciones en las que se aplican quimioterapias y estimar el retraso en la aparición de la recurrencia.

Es importante separar los objetivos individualmente porque nos permitirá desarrollar scripts individuales para cada objetivo y poder validar cada parte del proceso de manera independiente.

1.5. Estructura

- Estado del arte: En esta parte del trabajo primero se hará una revisión del estado del arte, donde incluiremos una exhaustiva revisión bibliográfica sobre modelos de cáncer sobre el estado de la modelización de cáncer.
- Análisis del problema: Aquí analizaremos el problema desde todas las perspectivas relevantes para el TFG. Primero haremos un análisis los datos que nos ha proporcionado el equipo médico, haremos un pequeño repaso de qué tecnologías y librerías podemos usar para desarrollar la solución, y haremos una breve introducción a la modelización matemática.
- Desarrollo de la solución propuesta: En este apartado relataremos los procesos por los que hemos tenido que pasar para obtener la solución.
- Pruebas: Mostraremos algunos resultados que da el programa para distintos tratamientos de quimioterapia. Estos consistirán en ir considerando distintos ciclos, y observar el retraso que nos indique el prototipo. No podremos realizar pruebas de validación sobre el prototipo debido a que el proyecto de investigación sigue en desarrollo.
- Conclusiones: En este apartado analizaremos si la solución obtenida hasta ahora ha sido capaz de satisfacer los objetivos que teníamos en mente, y las conclusiones a las que nos ha llevado la solución desarrollada. También plantearemos posibles mejoras. Finalmente, se analizarán los conocimientos obtenidos durante el desarrollo del TFG.

Capítulo 2. Estado del arte

2.1. Revisión bibliográfica

Buscando en el polibuscador, no he encontrado TFG que estuviesen relacionados con este tema, usando como palabras clave “bioinformática”, “modelo de cáncer” y similares. Lo único que he encontrado al respecto es un artículo sobre el cáncer de vejiga hecho en el Instituto de Matemática Multidisciplinar^[3].

Debido a esto evaluamos que es algo que no se ha explorado antes y que merece la pena explorar.

Posteriormente, como nos enseñaron en la asignatura de Bioinformática, miramos en el NCBI y los términos más reconocibles en el mundo de la medicina, como “HER2” y “cancer growth model”, para asegurarnos de esto también desde el punto de vista de ser un artículo de investigación, pero ninguno tenía el mismo enfoque que el nuestro, así que este tema estaba aparentemente libre en ambos casos.

Hemos hecho una revisión bibliográfica sobre el tema, y hemos dividido los trabajos encontrados entre los que estudian modelos de crecimiento de tumores, modelos de tumores con inmunoterapia, modelos de cáncer de mama y su interacción con el sistema inmune, modelos generales de cáncer con quimioterapia y finalmente de modelos de cáncer de mama con quimioterapia.

Antes de entrar en los detalles, queremos hacer una pequeña introducción sobre este tipo de modelización. Como sabemos, la modelización matemática pretende describir un fenómeno mediante expresiones matemáticas. La forma más común de estas expresiones suele ser las ecuaciones diferenciales o los sistemas de ecuaciones diferenciales. Desafortunadamente, solo un pequeño grupo de ecuaciones diferenciales se pueden resolver de forma exacta. La mayoría requieren de métodos numéricos para su resolución. Sin embargo, matemáticamente se puede realizar un estudio llamado análisis dinámico, que consiste en describir el comportamiento del sistema cuando se alcanza un cierto equilibrio. Este equilibrio se identifica con lo que ocurre a largo plazo. Así, pueden aparecer:

- Puntos de equilibrio: Puntos en los que se estabiliza el sistema.
- Ciclos límite: Situaciones estacionales que aparecen de forma estable.
- Bifurcaciones: Puntos alrededor de los cuales los comportamientos del sistema cambian sensiblemente.
- Comportamientos caóticos: Situaciones en las que el modelo proporciona resultados muy distintos a cambios aparentemente minúsculos del valor de sus parámetros, etc.

Existen diversas técnicas que permiten este estudio así como si esos comportamientos son estables o no.

El análisis dinámico es muy utilizado para el estudio del comportamiento de modelos complejos.

En las referencias [4][5] aparecen modelos de crecimiento de tumores y su idoneidad para representar su crecimiento.

En las referencias [6][7][8][9][10][11][12][13][14][15][16][17][18][19] aparecen modelos generales de tumores y su interacción con la respuesta inmune del individuo. En la mayoría de estos trabajos se realiza un análisis dinámico. En algunos se realizan simulaciones con datos recogidos de la literatura, si bien los datos y los parámetros de los modelos provienen de experimentos con ratas.

Las referencias [20][21][22][23][24][25] proponen modelos de la interacción entre el cáncer de mama y el sistema inmune, utilizando los enfoques mencionados anteriormente.

Hay pocas referencias donde se modelice el crecimiento de un cáncer junto con un tratamiento de quimioterapia. En [26] se proponen y estudian diversos modelos. En [27] se propone un modelo matemático que gobierna el crecimiento del cáncer a nivel de población celular con tratamientos combinados de inmunidad, vacunas y quimioterapia. Se realiza un análisis dinámico y simulaciones con datos de ratas y otros datos generales experimentales. En [28] el modelo propuesto es con retardo, con el que describen la ciclicidad en el crecimiento tumoral. Se realiza un análisis dinámico y unas simulaciones.

Tampoco hay muchas referencias que traten la modelización del cáncer de mama con tratamientos de quimioterapia. La estrategia fundamental de la quimioterapia es maximizar la erradicación del tumor dentro de los límites de toxicidad tolerable para el organismo. Para demostrar el uso de modelos matemáticos en el diseño de protocolos de tratamiento, en [29] se modeliza el efecto de la quimioterapia sobre la masa tumoral y se simulan varios protocolos. En el estudio se simularon tres posibles modos de crecimiento (exponencial, Gompertz y leyes de potencia). Los parámetros del modelo se estimaron mediante el procedimiento de mejor ajuste a partir de los datos clínicos de regresión tumoral tras tratamientos. Los parámetros estimados se utilizaron para simular varios regímenes que se emplean hoy en día en el tratamiento del cáncer de mama. Los resultados del estudio aportan información sobre la eficacia de la quimioterapia y pueden ayudar a diseñar mejores protocolos.

En [30] se estudia los efectos de la radioterapia y la quimioterapia combinadas como forma de tratar el cáncer de mama, y se incorporan al modelo distintos enfoques de tratamiento. Además, el modelo se ajusta a los datos de pacientes con cáncer de mama. Se determinan nuevas estrategias de tratamiento y se realiza un análisis dinámico.

En [31] se construye un sistema de ecuaciones diferenciales de orden fraccionario para analizar el crecimiento del cáncer de mama en el proceso de tratamiento inmuno-quimioterapéutico bajo algunos parámetros de control: dieta cetogénica, potenciador inmunológico y fármacos anticancerígenos.

2.2. Propuesta

La objetivo que tiene el proyecto aquí presentado es cuantificar el retraso en la aparición de recurrencias en cáncer de mama en mujeres en la que se administra quimioterapia preventiva tras extirparles el tumor. Una vez calibrado el modelo con la información de una paciente real antes de la resección, podemos proyectar a futuro el modelo para saber cuando puede presentarse la recurrencia. Así, podemos simular la administración de ciclos de quimioterapia y estimar el retraso que puede tener la paciente en la aparición de la recurrencia.

Debido a que los datos que nos ha proporcionado el equipo médico pueden tener errores, también buscamos tratar la incertidumbre del modelo en dos partes. Primero, tomamos los datos de la paciente y los expresamos como variables aleatorias Gaussianas, y segundo, hacemos que el modelo del equipo matemático nos devuelva funciones de densidad de probabilidad. Con estas dos cosas, podemos expresar con probabilidades los tamaños del tumor, y al hacer predicciones, poder estimar la probabilidad de que ocurra una recurrencia en un tiempo dado.

Nuestra propuesta abarca concretamente el ámbito informático, y colaboramos con un equipo matemático y un equipo médico. Esto es porque la solución que desarrollemos estará basada en el modelo que genere el equipo matemático, y que estará basado en la información que nos proporcione el equipo médico.

La realización de un modelo para predecir el cáncer de mama no es nuevo, pero nosotros lo hemos enfocado a poder retrasar la aparición de una recurrencia, y además queremos tratar la incertidumbre en el modelo para darle un valor añadido.

Capítulo 3. Análisis del problema

3.1. Datos disponibles

El equipo médico nos proporcionó los datos de una paciente. Solamente tratamos los datos de una paciente en concreto para la realización del prototipo. A continuación se muestra una tabla con los datos que utilizamos en el desarrollo de la solución final:

t (días)	T_{eco} (cm ³)	Ciclo	Tipo
0	0.046	-	-
15	-	-	-
23	-	1er	Docetaxel, Carboplatino, Trastuzumab
42	0.04	-	-
49	-	-	-
70	-	2do	Taxol, Pertuzumab, Trastuzumab
72	-	-	-
91	-	3er	Taxol, Pertuzumab, Trastuzumab
111	-	4to	Taxol, Pertuzumab, Trastuzumab
112	-	-	-
131	0.013	-	-
133	-	-	-
154	-	5to	Taxol, Pertuzumab, Trastuzumab
156	-	-	-
157	-	6to	Pertuzumab, Trastuzumab
233	Día de la resección		

Tabla 2: Tabla con los datos que nos facilitaron los médicos de una paciente

Donde:

- **t**: El paso temporal en días. Desde un $t_0=0$ donde inicialmente se detectó el tumor hasta el día de la resección.

- **T_{eco}** : Los tamaños del tumor de mama en cm³ obtenidos mediante ecografía. La recogida de estos datos contiene errores de medida, tanto por parte del médico como del dispositivo que toma las medidas, lo que justifica considerar la incertidumbre en el modelo.

- **Ciclo**: Número del ciclo de quimioterapia. Los tratamientos de quimioterapia se administran en ciclos que, según la literatura, suelen estar separados por 21 días.

- **Tipo**: Medicamento empleado en conjunto con el ciclo de quimioterapia. Los tratamientos de quimioterapia suelen acompañar los ciclos con medicamentos dependiendo de la estado del cáncer de mama.

3.2. Herramientas disponibles

Nos integramos en un grupo de investigación del Instituto Universitario de Matemática Multidisciplinar, donde trabajaban con Python, Mathematica y MATLAB.

También tenían una implementación del algoritmo PSO en Python, y nos podían asistir con la partes más matemáticas y estadísticas de la implementación. Esto, y el hecho de que ya teníamos experiencia de usar Python, hizo que la elección acabase siendo Python. Además, esto permitió que pudiésemos usar librerías más especializadas para esta clase de tareas como Scipy o Seaborn. A continuación hablaremos de las herramientas disponibles y justificaremos su elección.

Particle Swarm Optimization es un algoritmo de optimización no lineal basado en el comportamiento de las bandadas de pájaros. Es un algoritmo bio-inspirado que trata de optimizar una función de manera iterativa mejorando una solución candidata hasta que no pueda reducirse más el error. Resuelve el problema teniendo una población de soluciones candidatas, llamadas partículas, y moviendo estas partículas en el espacio de búsqueda de acuerdo a una fórmula que tiene en cuenta la posición y la velocidad de la partícula y de la mejor partícula (global best). El movimiento de cada partícula es influenciado por su mejor posición local conocida, y guiada hacia las mejores posiciones conocidas en el espacio de búsqueda, que se actualizan según se van encontrando mejores posiciones.

Mathematica⁴ es un programa que con librerías de distintas áreas que permiten, entre muchas cosas, optimizar funciones, representarlas en gráficas e implementar algoritmos. Esta es una de las herramientas que llegamos a usar debido a que el equipo matemático estaba muy familiarizado con ella y eran capaces de hacer breves demostraciones que ayudaban a que se entendiese mejor cuál era el objetivo de cada parámetro en el modelo.

MATLAB⁵ es un lenguaje de programación multiparadigma y un entorno de cálculo numérico desarrollado por MathWorks. MATLAB permite la manipulación de matrices, el trazado de funciones y datos, la implementación de algoritmos, la creación de interfaces de usuario y la interconexión con programas escritos en otros lenguajes.

Python⁶ es un lenguaje de alto nivel de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código, se utiliza para desarrollar aplicaciones de todo tipo. Se trata de un lenguaje de programación multiparadigma, ya que soporta parcialmente la orientación a objetos, programación imperativa y, en menor medida, programación funcional. Es un lenguaje interpretado, dinámico y multiplataforma. Es uno de los lenguajes de programación más populares y eso se refleja en la gran cantidad de librerías que tiene. Al final acabamos desarrollando el prototipo con Python, debido a que tenía todas las librerías que necesitábamos y una librería que implementaba el Particle Swarm Optimization desarrollada en el Instituto Universitario de Matemática Multidisciplinar, lo que ayudó mucho a poder desarrollar el prototipo.

4 <https://www.wolfram.com/mathematica/>

5 <https://www.mathworks.com/products/matlab.html>

6 <https://www.python.org/>

NumPy⁷ es una biblioteca para el lenguaje de programación Python que da soporte para crear vectores y matrices grandes multidimensionales, junto con una gran colección de funciones matemáticas de alto nivel para operar con ellas. NumPy es la librería que ayuda a que las matrices y sus operaciones que se realicen en Python sean eficientes, así que es una de las librerías que usamos para desarrollar el prototipo.

SciPy⁸ es una biblioteca libre y de código abierto para Python. Se compone de herramientas y algoritmos matemáticos. SciPy contiene módulos para optimización, álgebra lineal, integración, interpolación, funciones especiales, FFT, procesamiento de señales y de imagen, resolución numérica de ecuaciones diferenciales y otras tareas para la ciencia e ingeniería. Concretamente, durante el desarrollo del prototipo empleamos sus funciones de resolución de ecuaciones diferenciales y sus funciones para interactuar con las funciones de densidad de probabilidad de las variables aleatorias que acabaremos usando.

Matplotlib⁹ es una biblioteca para la generación de gráficos en dos dimensiones, a partir de datos contenidos en listas o arrays en el lenguaje de programación Python. Si no se usa directamente, sirve de núcleo para el resto de librerías gráficas y es la manera que tenemos de representar los datos para que el equipo matemático y el equipo médico los interprete.

Pandas¹⁰ es una librería de Python especializada en la manipulación y el análisis de datos. Ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales. Principalmente la usamos en el prototipo debido a que muchas de las librerías gráficas que empleamos funcionan mucho mejor con si se guardan los resultados en DataFrames, que es la estructura de datos que implementa Pandas.

Seaborn¹¹ es una librería de visualización de Python basado en matplotlib. Proporciona una librería de alto nivel para dibujar gráficas estadísticas mucho más visuales. Esta librería es la que nos permite representar los datos en conjunto con matplotlib.

Plotly¹² es una compañía que desarrolla herramientas de análisis de datos y herramientas de visualización. Nosotros usaremos su librería de visualización de datos, debido a que nos permite representar las gráficas de las funciones de densidad de probabilidad a lo largo del tiempo.

7 <https://numpy.org/>

8 <https://scipy.org/>

9 <https://matplotlib.org/>

10 <https://pandas.pydata.org/>

11 <https://seaborn.pydata.org/>

12 <https://plotly.com/python/>

GitLab¹³ es un servicio web de forja, control de versiones y DevOps basado en Git. Además de gestor de repositorios, el servicio ofrece también alojamiento de wikis y un sistema de seguimiento de errores, todo ello publicado bajo una licencia de código abierto. Se usó una suite de Gitlab disponible en el Instituto Universitario de Matemática Multidisciplinar para realizar el control de versiones.

Aunque la eficiencia algorítmica se mire cuando se tenga la solución desarrollada, durante el desarrollo se tuvo especial cuidado en usar *arrays* de Numpy siempre que se pudiera, ya que es una librería basada en C, y las operaciones que se encuentran en dicha librería son mucho más eficientes en tiempo y memoria debido a que no hacen las suposiciones sobre los tipos de datos que hace Python por defecto. También se tuvo cuidado de no hacer conversiones entre *arrays* de Numpy y listas de Python, debido a que son muy ineficientes.

De todas estas herramientas, usamos Mathematica cuando teníamos que hacer demostraciones rápidas sobre el comportamiento de las gráficas. MATLAB se descartó a favor de Python, Python es el lenguaje de programación en el que está escrito todo el prototipo, y sus librerías NumPy, SciPy, Pandas, Matplotlib, Seaborn y Plotly se usaron en distintas partes del prototipo para tratar los datos y poder representarlos gráficamente con la mayor facilidad posible.

MATLAB es uno de los lenguajes de programación que empleamos para el desarrollo del prototipo, pero que finalmente fue descartado a favor de Python debido a que nos resultaba más sencillo usar las librerías de Python para resolver el modelo descrito por ecuaciones diferenciales, y que el equipo matemático del Instituto Universitario de Matemática Multidisciplinar ya disponía de una implementación del algoritmo PSO. El equipo matemático también se aseguró de explicarnos todos los conceptos y técnicas matemáticas que hiciesen falta para el correcto desarrollo del prototipo.

Un aspecto a tener en cuenta es que la solución final se ha desarrollado en 3 scripts distintos debido no sólo a que nos permite diferenciar claramente los distintos pasos por los que tenemos que pasar para llegar a la solución que queremos, sino que al separarlos hacemos que no sea necesario volver a tener que ejecutar por los pasos anteriores. Esto es relevante porque la ejecución del algoritmo PSO para la calibración del modelo es la parte más costosa y difícil de estos tres scripts, lo que hace que no tengamos que ejecutar el calibrado más veces de las estrictamente necesarias.

Modelización matemática

La modelización matemática es una herramienta importante para describir, estudiar y predecir la evolución de procesos reales. Los modelos matemáticos se pueden describir por ecuaciones en diferencias o ecuaciones diferenciales, por redes, por programación lineal, etc. Dependiendo de la naturaleza del objeto de estudio, los

13 <https://about.gitlab.com/>

datos disponibles o el objetivo final de la investigación, los modelos se pueden expresar de una forma u otra.

En los últimos años, el tratamiento de la incertidumbre se ha convertido en una herramienta muy efectiva a la hora de estudiar fenómenos mediante modelos matemáticos. El motivo principal es porque incorporar incertidumbre en el modelo puede llevar a mejores descripciones de fenómenos reales, y en consecuencia a un estudio y predicción más acertados.

La incertidumbre del modelo normalmente se puede incorporar de dos formas distintas:

- La primera consiste en añadir aleatoriedad a las variables de entrada del modelo, donde en vez de ser valores deterministas, estos se consideran variables aleatorias. Esta es la manera en la que vamos a implementarlo en este TFG.
- La segunda consiste en incorporar un término, normalmente llamado “ruido blanco”, en la parte derecha de una ecuación diferencial. El “ruido blanco” engloba en un solo término matemático toda la posible incertidumbre que pueda haber en la formulación de un modelo, en los parámetros y en los datos.

Cuando la incertidumbre se añade a los parámetros del modelo mediante variables aleatorias, se asume que se conocen sus distribuciones de probabilidad y que siguen patrones estándar, o en otras palabras, se suelen considerar como variables aleatorias clásicas (Gaussianas, Uniformes, Betas, Gammas, entre otras).

Determinar distribuciones estadísticas que encajen a los parámetros aleatorios del modelo es un factor decisivo en el proceso de modelado, especialmente cuando se estudia un problema real. Cuando se describe un fenómeno, no es suficiente construir un modelo consistente, sino también considerar la distribución de probabilidad apropiada en los parámetros aleatorios del modelo.

3.3. Análisis del marco legal y ético

El equipo médico nos proporcionó los datos del tamaño de tumor, las fechas de las aplicaciones de los distintos ciclos de quimioterapia y el día de la resección de una paciente. Los datos están anonimizados, dado que son sensibles, y han sido tratados previamente por los médicos con sus respectivos análisis y restricciones.

Todas las librerías usadas están bajo el *BSD 3-Clause "New" or "Revised" License*^{14 15 16}, o el *MIT License*¹⁷, que ambos permiten usar la copia del software sin restricción, incluyendo los derechos a usar, copiar y distribuir.

14 <https://github.com/numpy/numpy/blob/main/LICENSE.txt>

15 <https://github.com/scipy/scipy/blob/main/LICENSE.txt>

16 <https://github.com/pandas-dev/pandas/blob/main/LICENSE>

3.4. Solución propuesta

La solución que se propone consiste en separar el prototipo en 3 partes. El funcionamiento de los scripts que desarrollaremos viene dado en la siguiente figura:

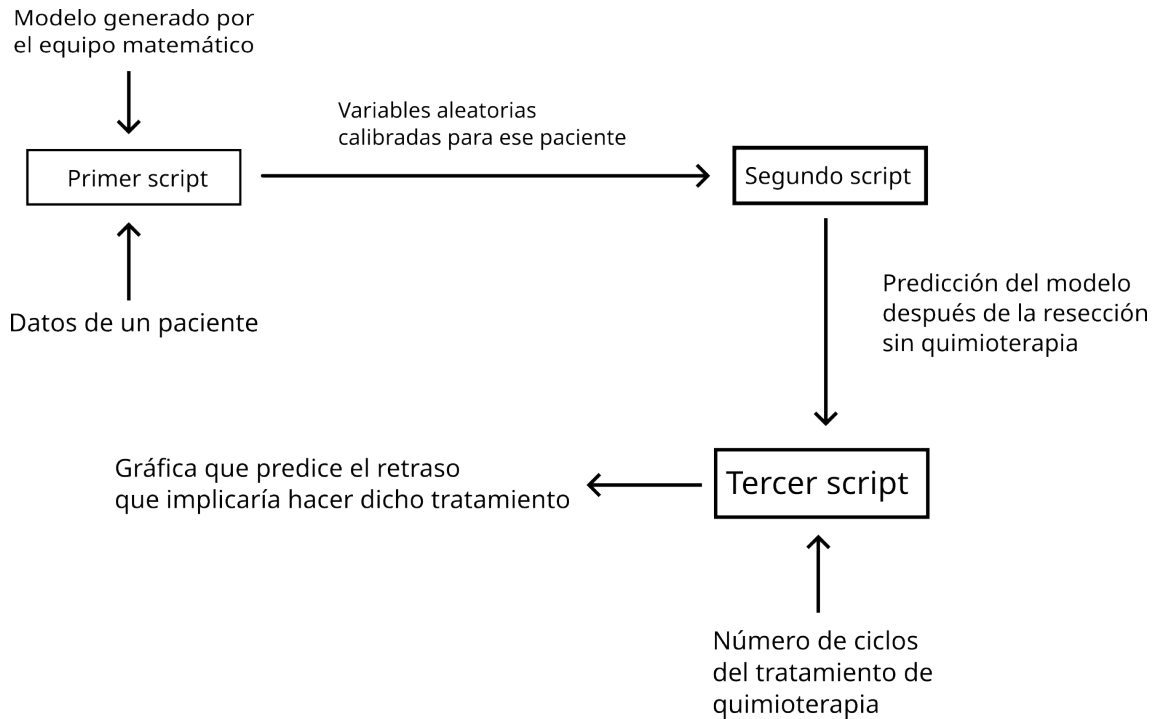


Figura 5: Diagrama que muestra el funcionamiento de los tres scripts desarrollados en conjunto

Donde:

- El primer script es el que necesitará inicialmente el modelo matemático a implementar y los datos de un paciente dado. El resto de scripts usarán el modelo que usemos aquí una vez se valide. Cuando hagamos la optimización con el PSO, el script nos devolverá las variables aleatorias calibradas para dicho paciente y nos mostrará una gráfica que vaya desde el inicio del tratamiento de quimioterapia hasta el día de la resección. El resultado estará expresado en forma de funciones de densidad de probabilidad para poder medir cómo influye en el modelo el error de los datos.

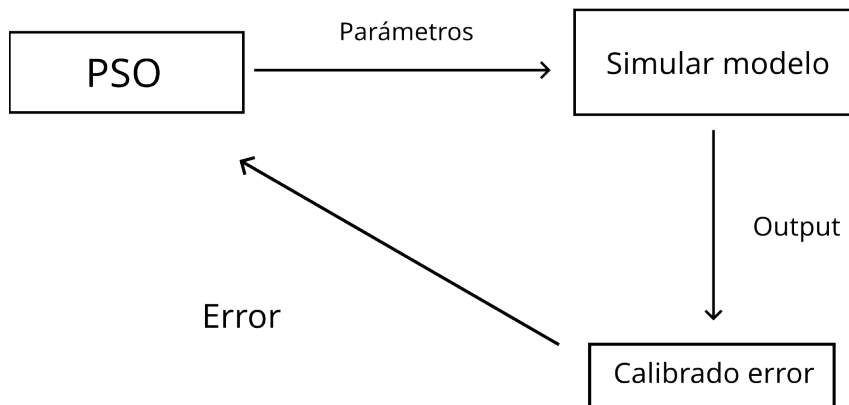


Figura 6: Diagrama que visualiza el proceso de optimización. (Primer script)

El diagrama de la Figura 6 muestra el funcionamiento que tendrá el primer script que estará compuesto por:

1. Una función *ajuste*($t0, k, a, b, t$), donde le pasaremos como argumentos las variables aleatorias del tamaño inicial del tumor $t0$, del crecimiento del tumor k , las dos que definen el comportamiento de la quimioterapia a y b , y un entero t indica el día en el que queremos la función de densidad de probabilidad, lo que nos permite ver qué probabilidad tiene el tumor de medir x en un instante t . Esta es la función que implementa el modelo que ha diseñado el equipo matemático.
 2. Una función *error*(*args*) a la que le pasaremos todos los argumentos que hagan falta en la función *ajuste*, y que se encargará de calcular el error cuadrático medio entre el output del modelo que devolverá la función *ajuste* y los datos.
 3. Una función principal *main*() donde llamaremos al algoritmo PSO, pasándole un rango de valores en el que se debe encontrar la solución, la función de *error*, y algunos valores que necesita el algoritmo para funcionar, como el número de ejecuciones y el tamaño del enjambre.
- El segundo script es el que, una vez se valide el modelo por el equipo matemático y el equipo médico, se encargará de calcular una predicción de los valores que tomará el tamaño del tumor después de la resección. Este script, además, devolverá una gráfica que será una función de densidad de probabilidad de los tiempos en los que el tamaño del tumor llegue al tamaño en el que se considera que ocurre una recurrencia.

Esto se obtendrá eliminando la función de *error*, cambiando la función de *ajuste* como nos indique el equipo matemático, y en la función principal llamamos a la función de *ajuste* con los parámetros calibrados en el primer script para todo instante t , y representando el resultado con una gráfica.

- El tercer script usa como base el segundo script, pero se le añaden las fechas de los ciclos de un hipotético tratamiento de quimioterapia y se comparan los resultados entre sí para poder evaluar un posible retraso en el crecimiento. Este script además nos mostrará la comparación entre las dos gráficas.

La implementación de la función *ajuste* en el segundo y el tercer script es la misma, simplemente cambia que al tercer script le pasamos los días en los que simulamos que ocurre un ciclo de quimioterapia, y que después medimos el tiempo que tarda el tumor en considerarse recurrencia en comparación a cuando no tenía ningún ciclo de quimioterapia aplicado.

Capítulo 4. Desarrollo de la solución propuesta

Para poder explicar el desarrollo de la solución final, tenemos que analizar el modelo que utilizaremos.

Sabemos que el tumor crece, y que la quimioterapia elimina células tumorales, así que estos comportamientos deben aparecer en la solución final. El modelo que propuso el equipo matemático es el siguiente:

$$\begin{cases} T'(t) = kT(t) - B \sum e^{-A(t-t_i)} H(t-t_i) T(t) \\ T(0) = T_0 \end{cases}$$

Figura 7: Modelo propuesto

Donde:

- **t**: Es el instante temporal en días.
- **T**: Es el tamaño del tumor en cm^3 .
- **k**: Es el parámetro a calibrar que influye al crecimiento del tumor.
- **B**: Es el parámetro a calibrar que influye en el efecto de la quimioterapia sobre el tumor.
- **A**: Es el parámetro a calibrar que influye sobre lo efectiva que es la quimioterapia respecto al tiempo transcurrido desde la aplicación del ciclo de quimioterapia.
- **H**: Función Heaviside con el objetivo de anular el término que corresponde a la quimioterapia cuando esta todavía no se ha aplicado. La función Heaviside se define como:

$$H(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 & \text{si } t \geq 0 \end{cases}$$

Figura 8: Definición de la función Heaviside

Esto nos deja con dos partes en el modelo. La que corresponde con el crecimiento del tumor, y la que corresponde con el efecto de la quimioterapia sobre el tumor, que son:

$$T'(t) = kT(t) \rightarrow \text{Crecimiento del tumor}$$

Figura 9: Fragmento del modelo completo que describe el crecimiento del tumor de mama

$$B \sum e^{A(t-t_i)} H(t-t_i) T(t) \rightarrow \text{Administración de la Quimioterapia}$$

Figura 10: Fragmento del modelo completo que describe el efecto de la quimioterapia sobre el crecimiento del tumor

La solución del modelo viene dada por la expresión:

$$T(t) = T_0 e^{kt + \frac{B}{A} \sum_{i=1}^N (e^{-A(t-t_i)} - 1) H(t-t_i)} \quad (cm^3)$$

Figura 11: Solución del modelo completo donde sustituimos T_0 , k , a y b por variables aleatorias para tratar la incertidumbre

Para poder tratar la incertidumbre, en el modelo completo consideramos las variables T_0 , k , a y b como variables aleatorias normales, de manera que vienen definidas por:

$$\begin{aligned} T_0 &\sim N(\mu_{T_0}, \sigma_{T_0}) & K &\sim N(\mu_K, \sigma_K) \\ A &\sim N(\mu_A, \sigma_A) & B &\sim N(\mu_B, \sigma_B) \end{aligned}$$

Figura 12: Parámetros del modelo expresados como variables aleatorias normales

Teniendo la solución de la Figura 11, teniendo en cuenta las distribuciones de los parámetros del modelo dadas en la Figura 12, utilizando la técnica de Random Variable Transformation (RVT), se puede obtener la función de densidad de probabilidad de la solución como Figura 13:

$$f_{T(t)}(y) = E_{A,B,K} \left[f_{T_0} \left(y e^{-kt - \frac{B}{A} \sum_{i=1}^N (e^{-A(t-t_i)} - 1) H(t-t_i)} \right) e^{-kt - \frac{B}{A} \sum_{i=1}^N (e^{-A(t-t_i)} - 1) H(t-t_i)} \right]$$

Figura 13: Fórmula que nos devuelve la función de densidad de probabilidad del tamaño del tumor respecto a un tiempo y espacio determinados

Random Variable Transformation^[32] es un método que nos permite obtener la distribución de densidad de probabilidad de un vector aleatorio V que sea producto de mapear otro vector aleatorio U cuya función de densidad de probabilidad sea conocida. En este caso, nos permite poder hallar la función de densidad de probabilidad del tamaño del tumor dado que conocemos las funciones de densidad de probabilidad de las variables aleatorias que la componen.

Al expresar el modelo de esta forma, hacemos que la implementación de la función en el código consista en sacar una cantidad de muestras de cada distribución truncada, y sacar la misma cantidad de resultados de esas muestras, que son las que aproximan la función de densidad de probabilidad del tamaño del tumor para ese instante t .

En cuanto a código, para poder manejar de manera más sencilla las variables aleatorias, creamos una clase *VariableAleatoria* que nos permitiese establecer la media y la varianza de una distribución, y con ellas poder acceder a una serie de funciones de Scipy respecto a esa media y esa varianza, lo que hace el código mucho más legible.

La clase *VariableAleatoria*¹⁸ funciona como un intermediario entre la variable aleatoria y Scipy, donde se establecen la media y la varianza de la distribución de antemano, y permite asignar la semilla para garantizar la reproducibilidad de la solución. Las funciones relevantes de esta función son:

- El constructor `__init__`: Función que construye el objeto *VariableAleatoria*, formado por un parámetro *media* y un parámetro *desviaciontipica*. Si no declaramos una *desviaciontipica* le asignamos una desviación típica por defecto a las variables del 10% de su media porque es lo que empleamos en la mayoría de los casos.
- La función *pdf*: Función que evalúa la función de densidad de probabilidad respecto a un valor o rango dado. Adicionalmente tiene un argumento *plot*, que si se asigna a *False*, simplemente se llama a la función *pdf* que nos proporciona Scipy con los atributos de la clase. Si se asigna el argumento *plot* a *True* se muestra un pequeño plot de dicha función de densidad de probabilidad, cuya función principal es de depuración.
- La función *rvs*: Función que genera una cantidad *size* determinada de muestras de la variable aleatoria. También tiene un argumento *plot* igual que el de *pdf*, e incluye un argumento *seed* para poder asignar una semilla y generar resultados reproducibles.
- La función *truncrvs*: Función que saca una cantidad determinada de variables aleatorias de una normal truncada. Esto nos sirve para asegurarnos de que las variables aleatorias sean siempre positivas para que los parámetros tengan sentido biológico. Por defecto trunca a un intervalo de confianza del 97.5%. También tiene argumento *plot* y argumento *seed* con la misma funcionalidad que en los casos anteriores.
- La función `__str__`: Es el equivalente a un `toString()`, pero en Python. Principalmente ayudó a depurar el funcionamiento de las *VariablesAleatorias*.

18 <https://rotondo.imm.upv.es/jovilta1/folder/-/blob/main/Nuevo/variablealeatoria.py>

4.1. El primer script

El primer script, "*grafica antes de la reseccion.py*"¹⁹ que es el encargado de calibrar los parámetros del modelo, esta formado por las siguientes funciones:

- La función *ajuste(t0, k, a, b, t)*, que es la función que devuelve la función de densidad de probabilidad del día *t*, dadas las variables aleatorias *t0, k, a, b*. Esta es la función que corresponde al modelo diseñado por el equipo matemático. La función contiene un bucle que corresponde al sumatorio del efecto de las quimioterapias en el instante *t*.
- La función *error(args)*, que es la que devuelve el error del modelo. Se llama a la función *ajuste* para los tiempos en los que tenemos datos, y se calcula el error cuadrático medio. Si el valor es negativo se aumenta el error para que evite estos valores.
- La función *main*, que es donde se declaran las variables que van a ser necesarias para el funcionamiento del algoritmo PSO, como el rango de valores que pueden tomar los parámetros del modelo, el número de ejecuciones, el tamaño del enjambre y la *seed* en caso de que queramos reproducir resultados previos. Se llama al algoritmo PSO con esas variables y la función de error. Posteriormente, calculamos las funciones de densidad de probabilidad desde *t=0* hasta el día de la resección pasando como argumentos a la función de ajuste las variables aleatorias de mínimo error y lo representamos gráficamente en tres dimensiones.

El resultado de la calibración devolvió los valores $K \sim N(0.003, 0.0001)$, $A \sim N(0.055299, 0.001)$ y $B \sim N(0.0278914, 0.001)$. La gráfica final resultante del primer script fue la que aparece en la Figura 14.

Donde:

- La *x* es el paso temporal en días, *y* es el tamaño de tumor y *z* es la frecuencia.
- Los puntos naranjas representan la media de los datos de tumor que tenemos convertidos a variables aleatorias, y es lo que usamos para hacer el calibrado de esta primera gráfica antes de la resección.

19 <https://rotondo.imm.upv.es/jovilta1/folder/-/blob/main/Nuevo/grafica%20antes%20de%20la%20reseccion.py>

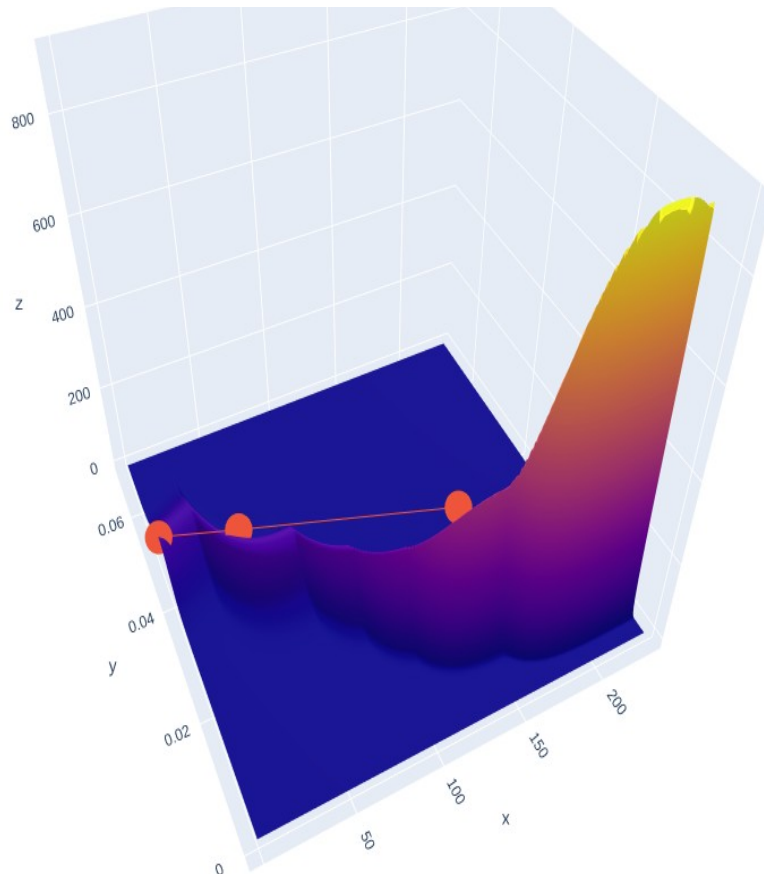


Figura 14: Gráfica que representa el crecimiento del tamaño del tumor como funciones de densidad de probabilidad. La x es el paso temporal en días, la y es el tamaño de tumor y z es la frecuencia.

4.2. El segundo script

Para el segundo script, “*grafica despues de la reseccion.py*”²⁰, tenemos que predecir el crecimiento del tumor después de la resección hasta que alcance un tamaño lo suficientemente grande como para considerarse recurrencia. Para conseguir esto, asumimos 3 hipótesis:

- Que las variables aleatorias K , A y B tienen el mismo valor antes y después de la resección.
- Que un tumor a partir de $T = 0.004 \text{ cm}^3$ se considera que hay recurrencia.
- De acuerdo a la literatura existen varios tamaños de tumor que se pueden quedar después de una resección, que son; $R_0 = T_0 \sim N(10^{-10}; 10^{-11})$, cuando el

²⁰ <https://rotondo.imm.upv.es/jovilta1/folder/-/blob/main/Nuevo/grafica%20despues%20de%20la%20reseccion.py>

tumor no se puede ver de manera microscópica, $R_1 = T_0 \sim N(10^{-8}; 10^{-9})$, cuando el tumor se puede ver de manera microscópica, y por último, $R_2: T_0 \sim N(10^{-6}; 10^{-7})$ ^[33], cuando el tumor se sigue viendo de manera macroscópica. Teniendo esto en cuenta, asumimos el peor de los casos $R_2 = T_0 \sim N(10^{-6}; 10^{-7})$.

Así que, partiendo del modelo con las variables transformadas, lo que tenemos que obtener es la función de densidad de probabilidad de los instantes de tiempo en los que T es igual a 0.004 cm^3 . Esto se expresa con la siguiente fórmula:

$$f_t(x) = E_{T_0} \left[f_K \left(\frac{1}{x} \log(T'/T_0) \right) \log(T'/T_0) \frac{1}{2} x^{-2} \right] (\text{cm}^3)$$

Figura 15: Fórmula que nos devuelve la función de densidad de probabilidad del tiempo en el que ocurre la recurrencia

donde T' es el tamaño 0.004 cm^3 . Respecto al código del segundo script, solamente usa la función de *ajuste*, donde K , A y B son las variables aleatorias calibradas, y el tamaño T_0 es R_2 . Esto devuelve la función de densidad de probabilidad de los tiempos t en días de cuándo pasa a considerarse que hay una recurrencia. La gráfica de dicha predicción es la siguiente:

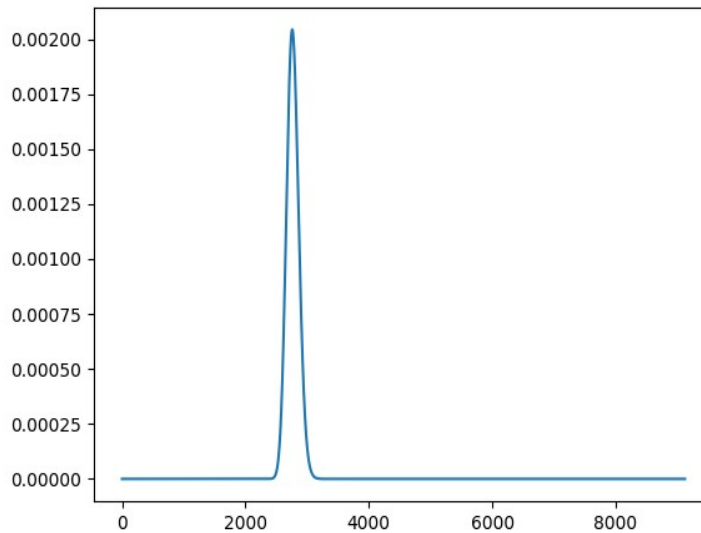


Figura 16: Gráfica que muestra la función de densidad de probabilidad de los tiempos en los que aparece una recurrencia donde el eje x es el paso temporal en días y y es la densidad de la función. Media = 2749 (días)

El eje x corresponde al paso temporal en días, y el eje y corresponde a la densidad de la función de densidad de probabilidad, indicándonos que tiempo es el más probable en el que ocurra la recurrencia. Lo representamos en una función de densidad para poder compararlo luego con los histogramas de las quimioterapias.

4.3. El tercer script

Y por último, en el script “*grafica despues de la reseccion con quimioterapia pso.py*”²¹ es donde juntaremos la gráfica de la Figura 16 con un histograma que represente el resultado de simular la aplicación de un tratamiento de quimioterapia.

En el tercer y último script, el modelo resuelto es el que nos permite comparar la función de densidad de probabilidad del tiempo con y sin quimioterapia. Dado el modelo de crecimiento propuesto, hay que muestrear las variables aleatorias y devolver cuándo llegan al tamaño que queremos, similar al script anterior. Para ello tenemos la solución del modelo completo, que corresponde a la Figura 11.

Aquí la función de error es la más simple de todas. Solamente comparamos la media que habíamos obtenido y marcado antes con la obtenida en el histograma obtenido administrando quimioterapia, y buscamos el tiempo para administrar quimioterapia de manera que retrase lo máximo posible la recurrencia.

En la función principal volvemos a establecer los límites superior e inferior que tomará el parámetro de en qué día administramos la quimio, y vamos ajustando el día. A continuación mostramos un ejemplo de cómo queda la gráfica, estableciendo como entrada un ciclo de quimioterapia después de un año:

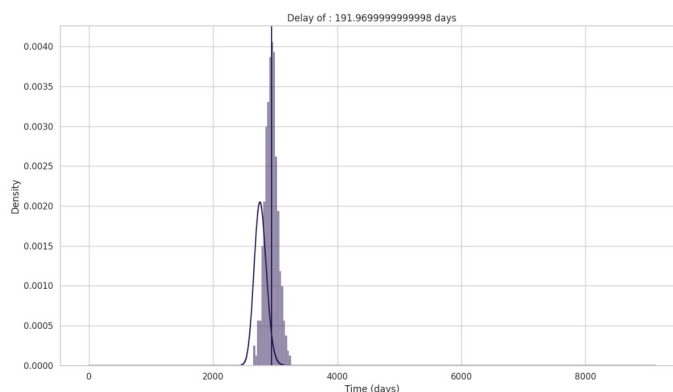


Figura 17: Gráfica que compara la gráfica de la Figura 16 con el histograma que obtenemos al aplicar una quimioterapia después de 1 año en dicha gráfica. El retraso en la recidiva es de media de 192 días.

21 <https://rotondo.imm.upv.es/jovilta1/folder/-/blob/main/Nuevo/grafica%20despues%20de%20la%20reseccion%20con%20quimioterapia%20pso.py>

Este último script es el que nos permite tomar las decisiones sobre cuando conviene poner quimioterapia, para retrasar al máximo la recidiva, y ese es el objetivo final de este TFG.

Capítulo 5. Pruebas

Inicialmente buscábamos optimizar la aplicación del tratamiento de quimioterapia, pero observamos que el retraso aumentaba únicamente cuando aumentábamos el número de ciclos del tratamiento. Esto significa que la solución a este problema no está acotada, y que si queríamos hacer pruebas, teníamos que considerar casos particulares en lugar de optimizar.

Acto seguido, comenzamos a probar distintos tratamientos. Primero probamos a aplicar una única quimioterapia en distintos días.

Recordamos que la **quimioterapia** es el tratamiento con medicamentos para interrumpir la formación de células cancerosas, ya sea mediante su destrucción o al impedir su multiplicación. La quimioterapia se administra por la boca, en inyección, por infusión o sobre la piel, según el tipo de cáncer y el estadio en que este se encuentra. Se administra sola o con otros tratamientos como cirugía, radioterapia o terapia biológica. Llamamos **ciclo de quimioterapia** al intervalo de tiempo comprendido entre dos sesiones de quimioterapia y que puede variar entre 21 a 28 días habitualmente. En ocasiones, dentro de un mismo ciclo se administran varias sesiones de quimioterapia.

De acuerdo a los datos que nos proporcionó el equipo médico, un tratamiento entero de quimioterapia consiste en 6 ciclos, y es lo que tomaremos como estándar cuando referenciamos un tratamiento completo de quimioterapia.

La primera gráfica [Figura 17] se obtuvo hacer una simulación con solo un ciclo de quimioterapia dejando un tiempo de reposo de un año después de la resección. Acto seguido hicimos pruebas con 1, 3 y 6 ciclos de quimioterapia, con distintos resultados. A continuación mostraremos las gráficas correspondientes:

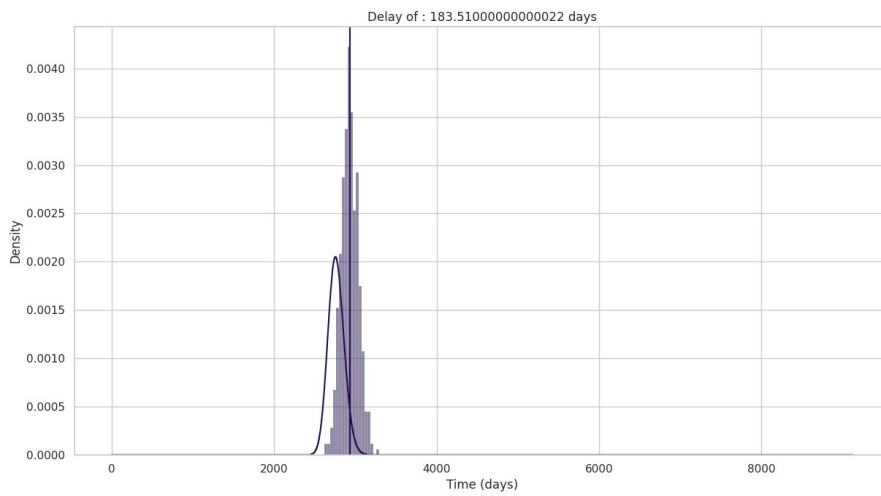


Figura 18: Gráfica que muestra el resultado de un ciclo de quimioterapia después de 1 año después de la resección. Muestra un retraso de 183 días.

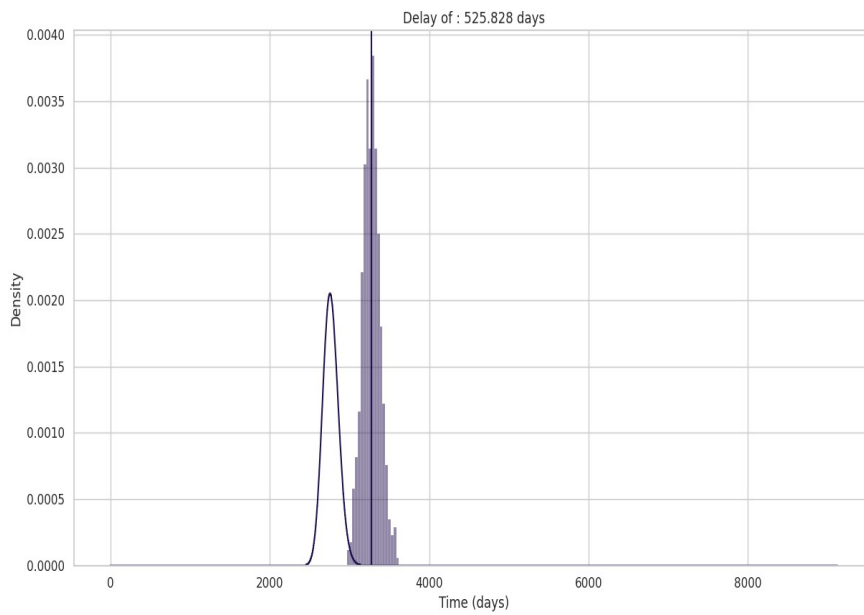


Figura 19: Gráfica que muestra el efecto de 3 ciclos de quimioterapia después de 1 año después de la resección. Muestra un retraso de 526 días.

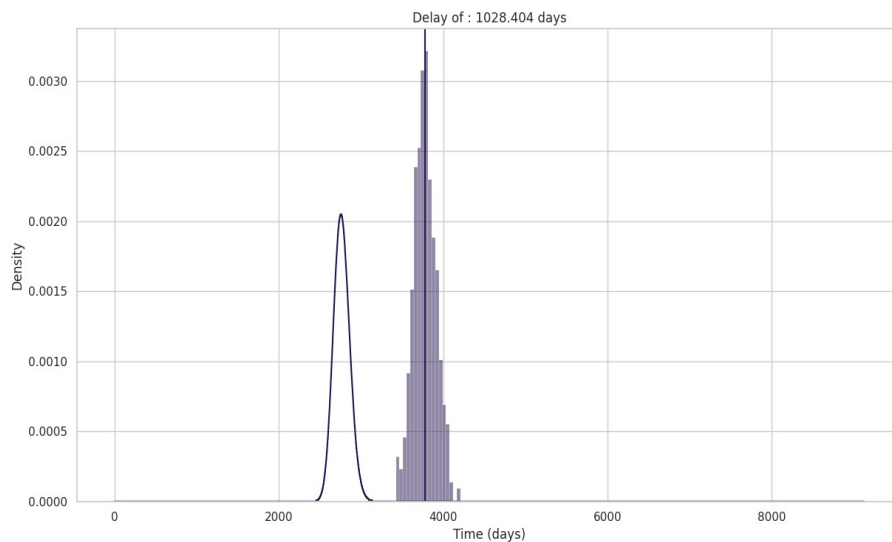


Figura 20: Gráfica que muestra el efecto de un tratamiento completo de 6 ciclos de quimioterapia que empieza 1 año después de la resección. Muestra un retraso de 1028 días.

Como resultado de esto se presentó una ponencia en el congreso MME&HB 2023²² titulada “Mathematical modeling of a HER2 breast cancer evolution after resection for a patient”.

22 <https://imm.webs.upv.es/jornadas/2023/home.html>

Capítulo 6. Conclusiones

A continuación vamos a comprobar que se han cumplido los objetivos establecidos:

- Inicialmente teníamos pensado poder desarrollar algo que funcionase más como un programa y que pudiese cubrir todos los casos. Sin embargo, cuando llegamos a fases más avanzadas del desarrollo y hacer cambios mucho más profundos sobre su funcionamiento, entendimos que era mucho más conveniente hacerlo en forma de scripts de Python, más modular, que pudiésemos cambiar con mucha más facilidad según iban cambiando los requisitos del proyecto. Teniendo esto en cuenta, el primer script, "*grafica antes de la reseccion.py*", es el que cumple el primer objetivo propuesto de representar gráficamente el crecimiento antes de la resección y devolver los parámetros del modelo calibrados.
- Le asignamos un tamaño inicial al tumor, pero es uno de los que se usan como referencia en el campo médico, concretamente R_2 . El resultado que se obtiene al simular la evolución del tamaño del tumor a partir de la resección usando los valores calibrados en el paso anterior, es concretamente una distribución de densidad de probabilidad de cuándo la recurrencia pasa el límite que le asignamos para considerar recurrencia. Con esto nos aseguramos de poder tratar la incertidumbre también después de la resección. El script "*grafica despues de la reseccion.py*" es es que cumple esta parte del objetivo.
- El script que cumple la última parte del objetivo de poder comparar el resultado del modelo cuando se le aplica quimioterapia es "*grafica despues de la reseccion con quimioterapia pso.py*".

La tecnología en la que más dominio se ha conseguido es sin duda Python. Gracias a la libertad de cómo poder desarrollar la solución hemos podido hacer uso de un gran número de librerías reconocidas en el mundo laboral, como Numpy, Scipy, Pandas y Seaborn. También hemos adquirido una base en conceptos de modelización, probabilidad para poder trabajar con matemáticos. Otro concepto es el algoritmo de optimización PSO, ya que en la carrera solamente se da optimización lineal.

Estos resultados que hemos obtenido son exploratorios, y no hemos podido validarlos todavía con los médicos. Sin embargo, las gráficas muestran una aparente relación entre el número de quimioterapias que se aplican al paciente y el retraso en la aparición de la recurrencia. Actualmente es un prototipo y conviene probar el modelo con más pacientes para afianzarlo.

Consideramos necesario probar el prototipo del programa con más pacientes para asegurarnos de que este comportamiento sea consistente en varios casos.

6.1.Relación del trabajo desarrollado con los estudios cursados

Este proyecto me interesó principalmente por una de las asignaturas optativas que cursé en los últimos cursos, que fue Bioinformática. De esa asignatura me resultaron de gran utilidad principalmente dos de las cosas que vimos, que fueron saber buscar información sobre el estado del arte, y la elaboración de un trabajo académico en forma de artículo científico que se realizó al final de la asignatura.

También caben destacar Algorítmica y Sistemas de Almacenamiento y Recuperación, una continuación de la otra, principalmente porque son las asignaturas en las que más usamos Python, y cuya práctica encadenada es lo más parecido que he hecho a este proyecto en la carrera respecto a tener en cuenta la estructura del código pensando más en los algoritmos que lo componen, más que en la estructura de las clases que se ve en otras asignaturas. Aun así sí que tuve que diseñar una clase para poder manipular y obtener información sobre las variables aleatorias sin tener que usar demasiado código redundante y poner en compromiso su legibilidad. Esto se refleja en la manera de elaborar el proyecto, en la que el equipo matemático nos proporciona un modelo ya generado, y tuvimos que estructurar el script de manera que lo que escriba en el script refleje la función.

Esto nos sirvió como excelente experiencia de aprendizaje y como una buena introducción al mundo de la investigación.

8. Bibliografía

- 1: An YY, Kim SH, Kang BJ, Differentiation of malignant and benign breast lesions: Added value of the qualitative analysis of breast lesions on diffusion-weighted imaging (DWI) using readout-segmented echo-planar imaging at 3.0 T, 2017, <https://doi.org/10.1371/journal.pone.0174681>
- 2: Cabrera Morales, C. M., Estudio comparativo de la amplificación de Her2/neu mediante FISH y PCR cuantitativa en tiempo real en tumores de mama, 2005, https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0378-48352005001000003
- 3: Cortés, J. C., López-Navarro, E., Moscardó-García, A., & Villanueva Micó, R. J., Calibración probabilística de un modelo de crecimiento del cáncer de vejiga teniendo en cuenta la incertidumbre de los datos., 2021-09-24, <http://hdl.handle.net/10251/180562>
- 4: John A. Adam, Nicola Bellomo, A Survey of Models for Tumor-Immune System Dynamics, Modeling and Simulation in Science, Engineering, & Technology, Jan/2011
- 5: Chivassa, Giorgia and Fornari, Chiara and Sirovichr, Roberta and Pennisi, Marzio and Beccuti, Marco and Cordero, Francesca, A mathematical model to study breast cancer growth, 2017, <https://doi.org/10.1109/BIBM.2017.8217874>
- 6: Shiferaw Feyissa and Sandip Banerjee, Delay-induced oscillatory dynamics in humoral mediated immune response with two time delays, 2013, <https://doi.org/10.1016/j.nonrwa.2012.05.001>
- 7: Marek Bodnar, Urszula Forys, Periodic dynamics in a model of immune system, 2018, <http://dx.doi.org/10.4064/AM-27-1-113-126>
- 8: Chaplain, Markand Matzavinos, Anastasios, Mathematical Modelling of Spatio-temporal Phenomena in Tumour Immunology, 2006
- 9: L.G. de Pillis and A. Radunskaya, A mathematical model of immune response to tumor invasion, 2003
- 10: Kirschner, D., Panetta, J. Modeling immunotherapy of the tumor – immune interaction, 1998
- 11: Kuznetsov, V.A., Makalkin, I.A., Taylor, M.A. et al, Nonlinear dynamics of immunogenic tumors: Parameter estimation and global bifurcation analysis, 1994
- 12: Frank Nani and H.I. Freedman, A mathematical model of cancer treatment by immunotherapy, 2000, [https://doi.org/10.1016/S0025-5564\(99\)00058-9](https://doi.org/10.1016/S0025-5564(99)00058-9)
- 13: A.S. Bratus and E. Fimmel and Y. Todorov and Y.S. Semenov and F. Nürnberg, On strategies on a mathematical model for leukemia therapy, 2012, <https://doi.org/10.1016/j.nonrwa.2011.02.027>

- 14: Alberto d'Onofrio and Francesca Gatti and Paola Cerrai and Luca Freschi, Delay-induced oscillatory dynamics of tumour-immune system interaction, 2010, <https://doi.org/10.1016/j.mcm.2009.11.005>
- 15: Denise E. Kirschner , Alexei Tsygvintsev, On the global dynamics of a model for tumor immunotherapy, 2009, <https://doi.org/10.3934/mbe.2009.6.573>
- 16: Shigui Ruan, Nonlinear dynamics in tumor-immune system interaction models with delays , 2021, <https://doi.org/10.3934/dcdsb.2020282>
- 17: Fuat Gurcan and Senol Kartal and Ilhan Ozturk and Fatma Bozkurt, Stability and bifurcation analysis of a mathematical model for tumor-immune interaction with piecewise constant arguments of delay, 2014, <https://doi.org/10.1016/j.chaos.2014.08.001>
- 18: Monika Joanna Piotrowska, An immune system-tumour interactions model with discrete time delay: Model analysis and validation, 2016, <https://doi.org/10.1016/j.cnsns.2015.10.022>
- 19: Andrei Korobeinikov and Konstantin E Starkov and Paul A Valle, Modeling cancer evolution: evolutionary escape under immune system control, 2017, <https://dx.doi.org/10.1088/1742-6596/811/1/012004>
- 20: Chipso Mufudza, Walter Sorofa, Edward T. Chiyaka, Assessing the Effects of Estrogen on the Dynamics of Breast Cancer, 2012, <https://doi.org/10.1155/2012/473572>
- 21: Angela M Jarrett, Meghan J Bloom, Wesley Godfrey, Anum K Syed, David A Ekrut, Lauren I Ehrlich, Thomas E Yankeelov, Anna G Sorace, Mathematical modelling of trastuzumab-induced immune response in an in vivo murine model of HER2+ breast cancer, 2019, <https://doi.org/10.1093/imammb/dqy014>
- 22: Oke, Segun Isaac and Matadi, Maba Boniface and Xulu, Sibusiso Southwell, Optimal Control Analysis of a Mathematical Model for Breast Cancer, 2018, <https://www.mdpi.com/2297-8747/23/2/21>
- 23: OPhir Nave and Miriam Elbaz and Svetlana Bunimovich-Mendrazitsky, Analysis of a breast cancer mathematical model by a new method to find an optimal protocol for HER2-positive cancer, 2020, <https://doi.org/10.1016/j.biosystems.2020.104191>
- 24: Xin Liand D. Thirumalai, A mathematical model for phenotypic heterogeneity in breast cancer with implications for therapeutic strategies, 2022, <https://doi.org/10.1098/rsif.2021.0803>
- 25: Mehdizadeh R, Shariatpanahi SP, Goliaei B, Peyvandi S, Rüegg C., Dormant Tumor Cell Vaccination: A Mathematical Model of Immunological Dormancy in Triple-Negative Breast Cancer, 2021 Jan 11, <https://doi.org/10.3390/cancers13020245>

- 26: Martin Eisen, Mathematical Models in Cell Biology and Cancer Chemotherapy, 1979
- 27: L.G. de Pillis and W. Gu and A.E. Radunskaya, Mixed immunotherapy and chemotherapy of tumors: modeling, applications and biological interpretations, 2006, <https://doi.org/10.1016/j.jtbi.2005.06.037>
- 28: Krishnapriya, P., Pitchaimani, M., Optimal control of mixed immunotherapy and chemotherapy of tumours with discrete delay, 2015
- 29: Eliezer Shochat, Dvora Hart, Zvia Agur, Using computer simulations for evaluating the efficacy of breast cancer chemotherapy protocols, 2002,
- 30: Sara B. Mkango, Nyimvua Shaban, Eunice Mureithi, Twalib Ngoma, Dynamics of Breast Cancer under Different Rates of Chemoradiotherapy, 2019, <https://doi.org/10.1155/2019/5216346>
- 31: Ali Yousef, Fatma Bozkurt, Thabet Abdeljawad, Mathematical modeling of the immune-chemotherapeutic treatment of breast cancer under some control parameters, 2020, <https://doi.org/10.1186/s13662-020-03151-5>
- 32: T. T. Song, Random Differential Equations in Science and engineering, 1973
- 33: Hermanek, Paul and Wittekind, Christian, Residual tumor (R) classification and prognosis, 1994, <https://doi.org/10.1002/ssu.2980100105>

Anexo ODS

OBJETIVOS DE DESARROLLO SOSTENIBLE

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS)

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza.				X
ODS 2. Hambre cero.				X
ODS 3. Salud y bienestar.	X			
ODS 4. Educación de calidad.				X
ODS 5. Igualdad de género.				X
ODS 6. Agua limpia y saneamiento.				X
ODS 7. Energía asequible y no contaminante.				X
ODS 8. Trabajo decente y crecimiento económico.				X
ODS 9. Industria, innovación e infraestructuras.			X	
ODS 10. Reducción de las desigualdades.				X
ODS 11. Ciudades y comunidades sostenibles.				X
ODS 12. Producción y consumo responsables.				X
ODS 13. Acción por el clima.				X
ODS 14. Vida submarina.				X
ODS 15. Vida de ecosistemas terrestres.				X
ODS 16. Paz, justicia e instituciones sólidas.				X
ODS 17. Alianzas para lograr objetivos.				X

El ODS con el que más se relaciona este TFG es el de Salud y bienestar, dado que es un TFG que busca aplicarse en medicina y salud.

El objetivo del TFG finalmente es hacer un script que nos pueda mostrar cuánto podemos llegar a retrasar la recurrencia de un cáncer de mama aplicando ciclos de quimioterapia a un paciente dado. Para poder hacer esto hemos tenido que colaborar con médicos del servicio de oncología. Esto sitúa al proyecto en un ámbito médico aparte de informático y matemático, y por eso consideramos que se relaciona con el ODS mencionado anteriormente.

Durante la realización del trabajo hemos tenido reuniones con distintos profesionales médicos, y siempre se ha tenido en cuenta la calidad de vida del paciente a la hora de establecer los ciclos de quimioterapia que se le pudiesen aplicar, manteniéndolos lo más cercano a cómo se aplican en terapias normales. Así garantizamos que no estamos proponiendo nada demasiado fuera de los tratamientos que ya se aplican. Debido a esto, también consideramos que el bienestar del paciente ha sido algo que hemos tenido en cuenta durante la realización de este TFG.

Hemos considerado también el ODS de Industria, innovación e infraestructuras, aunque en menor medida, principalmente por el apartado de innovación. Este TFG se podría considerar innovador al introducirse en un área no explorada previamente.

Respecto a otros ODS, la temática de este trabajo es bastante centrada y no se presta mucho a objetivos más sociales. Tampoco lo hemos podido relacionar con ODS económicos debido a que todavía es un prototipo, y fue desarrollado con el propósito de investigar.