

ESTIMACIÓN DE PARÁMETROS PARA EL AJUSTE DE
UNA DINÁMICA DE LA TUBERCULOSIS EN LA
CIUDAD DE CALI

JOHAN SEBASTIAN LEAL TORO

UNIVERSIDAD DEL VALLE
FACULTAD DE CIENCIAS NATURALES Y EXACTAS
DEPARTAMENTO DE MATEMÁTICAS
SANTIAGO DE CALI

2024

ESTIMACIÓN DE PARÁMETROS PARA EL AJUSTE DE UNA DINÁMICA DE LA TUBERCULOSIS EN LA CIUDAD DE CALI

JOHAN SEBASTIAN LEAL TORO

*Proyecto de grado presentado como requisito para optar al título
de Matemático*

Directora:

JUDDY HELIANA ARIAS CASTRO, Dra. C.

Co-Director:

EDWIN BARRIOS RIVERA, Dr. C.

UNIVERSIDAD DEL VALLE
FACULTAD DE CIENCIAS NATURALES Y EXACTAS
DEPARTAMENTO DE MATEMÁTICAS
SANTIAGO DE CALI

2024

Índice general

Introducción	1
1. Preliminares	3
1.1. La Tuberculosis	3
1.2. Modelamiento matemático	4
1.3. Estimación de parámetros	5
1.3.1. Método de mínimos cuadrados (MC)	6
1.3.2. Algoritmos genéticos (AG)	7
1.4. Datos para el modelo	8
2. Modelo de transmisión de la Tuberculosis (TB)	10
2.1. Descripción del modelo	10
2.2. Número básico de reproducción ($\mathcal{R}_0$)	13
2.3. Sensibilidad del $\mathcal{R}_0$	13
2.4. Condiciones Iniciales	15
2.5. Modelo de TB normalizado	15
3. Programas y Resultados	17
3.1. Rutina de <i>lsqcurvefit</i> de Matlab	18
3.2. Algoritmo genético	19
3.3. Resultados	20
3.4. Discusión de resultados	25
4. Conclusiones	27
Bibliografía	29

Índice de figuras

1.1.	Diagrama de un modelo tipo SIR.	5
1.2.	Genes como estructura de cadena.	8
1.3.	Ejemplo de crossover.	8
2.1.	Diagrama del modelo de la dinámica de la tuberculosis dada en [5].	11
3.1.	Ajuste realizado por <i>lsqcurvefit</i> y solución de la ecuación de tratados en el sistema (2.2).	22
3.2.	Soluciones del sistema (2.2) con los parámetros encontrados por <i>lsqcurvefit</i> . .	22
3.3.	Errores cuadráticos de <i>lsqcurvefit</i>	23
3.4.	Ajuste realizado por el algoritmo genético y solución de la ecuación de tratados en el sistema (2.2).	24
3.5.	Soluciones del sistema (2.2) con los parámetros encontrados por el algoritmo genético.	24
3.6.	Errores cuadráticos del algoritmo genético.	25

Índice de cuadros

1.1.	Parámetros para el modelo tipo SIR representado en la Figura 1.1.	5
2.1.	Descripción de los parámetros del modelo de TB dado en [5].	12
2.2.	Rangos para los parámetros del modelo presentados en [5] ajustados a los datos de Cali en meses.	14
2.3.	Índice de sensibilidad del $\mathcal{R}_0$ respecto a cada parámetro.	14
2.4.	Conjunto de parámetros a estimar.	15
2.5.	Condiciones iniciales del sistema.	15
3.1.	Rangos de los parámetros a estimar en el modelo normalizado.	17
3.2.	Condiciones iniciales del sistema normalizado.	18
3.3.	Vectores de condiciones iniciales para los parámetros.	21
3.4.	Tabla de resultados para las simulaciones de la rutina <i>lsqcurvefit</i> con las condiciones iniciales de los parámetros dadas en la Tabla 3.3.	23
3.5.	Tabla de resultados para las simulaciones del algoritmo genético (AG) con las condiciones iniciales de los parámetros dadas en la Tabla 3.3.	25

Resumen

La Tuberculosis (TB) es una enfermedad infecciosa que se transmite de manera directa, la cual ha venido afectando la población desde varios siglos atrás; por ello, en epidemiología se han realizado gran variedad de modelos matemáticos para el estudio de la dinámica de esta enfermedad. En este documento, se revisa el modelo en ecuaciones diferenciales ordinarias para la dinámica de la transmisión de la enfermedad de tuberculosis presentado en [5], y usando datos de personas en tratamiento en la ciudad de Santiago de Cali, se estiman algunos de los parámetros del modelo utilizando los softwares de programación **MATLAB** y **Wolfram Mathematica**, mediante el método de estimación de parámetros de mínimos cuadrados no lineales y con dos algoritmos de optimización distintos. La estimación de estos parámetros, como en todo modelo epidemiológico, es bastante importante ya que nos permite usar el modelo para analizar el comportamiento de la enfermedad dentro de la población estudiada y establecer medidas de control dependiendo de los valores encontrados. Los métodos de estimación de parámetros aquí estudiados son el algoritmo genético y el método de mínimos cuadrados no lineales y, la selección de los parámetros a estimar se realiza a partir del análisis de sensibilidad del número básico de reproducción $\mathcal{R}_0$.

Palabras clave: tuberculosis, modelamiento, mínimos cuadrados, algoritmo genético, tratados.

Introducción

Uno de los tantos problemas de salud pública está constituido por la Tuberculosis (TB). Según información de la Organización Panamericana de Salud (OPS), la Tuberculosis es una enfermedad infecciosa que suele afectar los pulmones entre otras partes del cuerpo. Esta enfermedad puede desarrollarse (tuberculosis activa) o no (tuberculosis latente) dependiendo del sistema inmune de cada persona y es causada por una bacteria llamada *Mycrobacterium tuberculosis*, que se transmite de persona a persona a través de gotitas o aerosoles suspendidos en el aire secretadas por una persona con tuberculosis activa al toser, hablar o estornudar [1]. Esta forma de transmisión ha puesto en riesgo la vida de un gran número de personas, desde niños hasta adultos mayores y ha representado un gran problema para las personas con afectaciones en su sistema inmune como el VIH, causas por malnutrición, diabetes, entre otros; pues al estar expuestos a la enfermedad, el desarrollo de esta es mas rápido. Siendo la TB una enfermedad que ha afectado la salud pública desde siglos atrás, según la Organización Mundial de la Salud (OMS), se calcula que actualmente una cuarta parte de la población mundial está infectada por el bacilo de la tuberculosis pero aún no han enfermado ni pueden transmitir la infección, es decir que se encuentran en estado de latencia [2].

En ciencias e ingeniería, el modelamiento matemático es una herramienta frecuentemente utilizada para el estudio de fenómenos físicos, químicos, biológicos, etc. Un modelo matemático utiliza fórmulas matemáticas para representar de manera simplificada y objetiva el fenómeno que se quiera estudiar; dichas fórmulas se plantean a partir del análisis del comportamiento del fenómeno. Una característica esencial de cualquier modelo son los parámetros utilizados; en el caso epidemiológico, se pueden considerar, por ejemplo: tasas de mortalidad, tasas de infecciosidad, probabilidades de contagio, tasas de crecimiento y decrecimiento, tasas de recuperación, entre otros. Algunos de estos parámetros pueden estimarse a partir de datos recolectados en cierto tiempo y con ayuda de los métodos de estimación de parámetros existentes como Mínimos Cuadrados, Máxima Verosimilitud o Estimación por cuantiles, y proporcionan información que podría ayudar a realizar predicciones, establecer límites y realizar inferencias sobre el fenómeno estudiado para una población en particular.

Desarrollar un modelo matemático significa obtener leyes que relacionen las variables involucradas en el fenómeno estudiado, donde cada una de ellas es representativa de las diferentes etapas o procesos que forman parte del mismo [14]. En epidemiología, un modelo matemático considera subdivisiones en la población y analiza el desarrollo de una enfermedad en un

tiempo determinado de manera local, pues dependiendo del entorno en que se encuentre la población estudiada, se pueden encontrar factores que afecten los datos con que se trabaja y por ende el resultado de cualquier estimación de parámetros realizada.

La estimación de parámetros juega entonces un papel importante para poder analizar la dinámica de un proceso que se ha modelado matemáticamente, pues los parámetros calculados deben ajustarse a los datos obtenidos para que el modelo pueda aproximarse al comportamiento real del fenómeno que se observa. Sin embargo, el método de estimación de parámetros empleado debe escogerse detenidamente, pues algunos optimizan más el problema del ajuste que otros, dependiendo de las restricciones que se le den y, la cantidad y tipo de datos con que se quiera trabajar.

A través de los años, se han desarrollado cierta cantidad de modelos matemáticos que han ayudado a estudiar la dinámica de la tuberculosis en diferentes lugares y con diferentes parámetros considerados. Por ejemplo, en [7], se presenta un modelo de la TB donde se introduce un grupo en la población, denominados latentes de bajo riesgo, para proponer estrategias óptimas de prevención y control de la enfermedad con ayuda de la teoría del control óptimo. Otro modelo de TB puede encontrarse en [19], en el cual se estudia el impacto de la vacunación y el tratamiento en la dinámica de transmisión de la enfermedad. En [18], se investigan dos modelos diferentes, uno que tiene en cuenta el fracaso en el tratamiento de la enfermedad y otro donde se dividen las personas latentes en dos subgrupos dependiendo del tipo de progreso, rápido o lento de la enfermedad. Es claro entonces, que dependiendo de lo que se quiera estudiar de la enfermedad, los modelos satisfacen características diferentes; el modelo planteado en [5], por ejemplo, estudia la dinámica de esta enfermedad considerando a las personas que abandonan el tratamiento volviendo nuevamente a un estado de latencia o de infecciosidad.

Para fines de este trabajo de grado, se considera el modelo con ecuaciones diferenciales ordinarias planteado en [5], para el cual se tiene como objetivo estimar algunos de sus parámetros utilizando los datos de personas en tratamiento de tuberculosis en el período 2015||2019 de la ciudad de Santiago de Cali reportados por la Secretaría Municipal de Salud Pública.

A lo largo de este documento, se encuentran dos algoritmos diferentes para la estimación de parámetros a partir del método de mínimos cuadrados no lineales: el método de Levenberg-Marquardt aplicado con una rutina optimizada de Matlab y el algoritmo genético, los cuales se ponen en práctica con el fin de estimar algunos de los parámetros del modelo estudiado y ayudan a comprender algunas de las características que los diferencian. Así pues, el orden en que se desarrolla este documento es el siguiente: inicialmente se da una introducción de la enfermedad de tuberculosis y la teoría correspondiente de los algoritmos para la estimación de parámetros utilizados; seguidamente, se presenta la descripción del modelo y la ecuación del número básico de reproducción $\mathcal{R}_0$ junto con sus índices de sensibilidad. Por último, se muestran los algoritmos de programación de cada método junto con los resultados obtenidos para los parámetros estimados y las gráficas del ajuste realizado.

Capítulo 1

Preliminares

Para dar inicio a este documento de trabajo de grado, es necesario tener en cuenta algunos conceptos como son: ¿qué es la enfermedad de tuberculosis?, ¿cómo se transmite?, la importancia del modelamiento matemático y sobre todo los métodos y datos utilizados para llevar a cabo el objetivo de estimación de parámetros. En esta sección, se abordarán estos temas.

1.1. La Tuberculosis

La tuberculosis (TB) es una enfermedad infecciosa causada por una bacteria conocida como *Mycobacterium tuberculosis* o bacilo de Koch [3]. Ésta suele atacar los pulmones, aunque también puede afectar otras partes del cuerpo como el cerebro o las vértebras, y puede desarrollarse en cada persona (TB activa) o no (TB latente) dependiendo de su sistema inmune, pues una persona con afectaciones en este, ya sea por enfermedades como la diabetes, el virus de inmunodeficiencia humana (VIH), u otras causas, es más propensa a desarrollar la enfermedad de tuberculosis de manera activa.

Su forma de transmisión es de manera directa por medio de gotitas o aerosoles suspendidos en el aire que son secretadas por una persona con TB pulmonar activa al toser, hablar o estornudar. Esta misma puede ser tratada con diferentes tipos de medicamentos, algunos de estos con una actividad bactericida los cuales se encargan de eliminar las bacterias de crecimiento continuo que son las que transmiten la enfermedad y otros con una actividad esterilizante que se encargan de eliminar los bacilos de TB en el cuerpo causantes de las recaídas [4].

La Organización Mundial de la Salud (OMS) afirma que alrededor de una cuarta parte de la población mundial se ha infectado con el bacilo de la tuberculosis y entre el 5 % y el 10 % de estas personas termina desarrollando la enfermedad, además las personas infectadas que aún no la han desarrollado, no pueden transmitirla [2]. En Colombia, según lo reportado por el Instituto Nacional de Salud, entre los años 2013 y 2018, la tuberculosis presentó un aumento constante en su tasa de incidencia desde los 23,0 hasta 26,9 casos por cada 100.000

habitantes, siendo una de las primeras causas de morbilidad en el país [10].

La TB es una enfermedad que puede ser mortal y algunos de sus principales síntomas son fiebre, tos prolongada (algunas veces con sangre), dolor en el pecho, cansancio, sudoración nocturna y pérdida de apetito. Esta enfermedad tiene varias formas de diagnosticarse, algunas de estas son la baciloscopia (examen de esputo o flema), prueba de sangre y prueba cutánea.

1.2. Modelamiento matemático

El modelamiento matemático ha sido una herramienta de gran importancia a la hora de analizar algún fenómeno. Esta herramienta se centra en representar de manera sintetizada y a través de ecuaciones el problema que se quiere estudiar. En [17], se define un modelo matemático como algo que usa el lenguaje de las matemáticas para producir una descripción refinada y precisa de un sistema de objetos y que, además, todos estos modelos están de cierta forma “errados”, pues cada uno de ellos siempre va a depender de qué tan completo sea, es decir, qué tantas características del objeto se tengan en consideración.

En algunos campos, como la epidemiología para analizar el comportamiento de una enfermedad a partir de una o un sistema de ecuaciones, es necesario establecer parámetros como por ejemplo tasas de mortalidad, natalidad o transmisión de la enfermedad; para esto, es necesario dividir la población en ciertos grupos teniendo en cuenta cómo se desarrolla la enfermedad, si es relevante o no considerar el periodo de latencia, si tiene tratamiento, si hay una recuperación parcial o total luego del tratamiento, entre otras condiciones. Para estos casos, surgen los llamados modelos compartimentales, pues nos permiten describir y estudiar la dinámica de un individuo u objeto que puede moverse entre las diferentes compartimientos. Volviendo al caso epidemiológico, y como se muestra en [8], los modelos compartimentales clásicos son SI, SIS y SIR. Estos parten de una división de la población considerada en tres clases que varían con el tiempo t :

- **Susceptibles (S):** en este grupo se encuentran los individuos que no han sido expuestos o no han interactuado con la enfermedad.
- **Infectados (I):** son los individuos con la enfermedad activa.
- **Recuperados (R):** en este grupo se encuentran quienes se hayan recuperado completamente de la enfermedad.

Si la recuperación no confiere inmunidad permanente, nos encontramos ante un modelo SIS, en el que los individuos se trasladan de la clase de susceptibles a la de infectados y al recuperarse vuelven a ser susceptibles, como lo es el caso de un gran número de enfermedades de transmisión sexual. Si los individuos que se recuperan son inmunes, entonces estamos ante un modelo SIR, donde caben modelos de enfermedades como la influenza (gripe). Si los individuos no se recuperan, estamos ante un modelo SI, tal es el caso del VIH.

Para una mejor comprensión de los modelos compartimentales, se realizan diagramas que

sirven para visualizar el comportamiento de la transición de individuos entre un grupo y otro, para lo cual es necesario establecer ciertas cantidades que servirán para identificar tales tasas de transición, estas cantidades son llamados parámetros del modelo. En la Figura 1.1, se presenta un diagrama para un modelo del tipo SIR y sus respectivos parámetros, sin considerar la muerte por la enfermedad, se encuentran descritos en la Tabla 1.1.

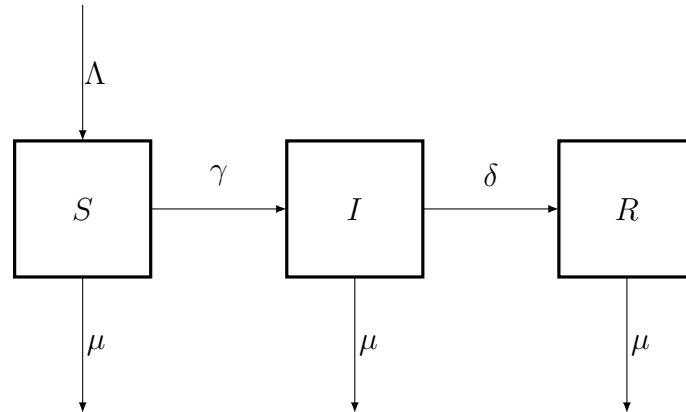


Figura 1.1: Diagrama de un modelo tipo SIR.

Parámetro	Descripción
Λ	Reclutamiento de individuos humanos
μ	Tasa de mortalidad natural
γ	Tasa de infecciosidad
δ	Tasa de recuperación

Tabla 1.1: Parámetros para el modelo tipo SIR representado en la Figura 1.1.

1.3. Estimación de parámetros

En ciencias e ingeniería, la estimación de parámetros es una herramienta importante que permite analizar una dinámica o problema modelado matemáticamente, con datos obtenidos a través de encuestas, experimentos y en general cualquier muestra estadística. Esto, por ejemplo puede verse en la experimentación física al momento de calcular la gravedad, el coeficiente de fricción, resistividad o índice de refracción de algún material, entre otros; pues dichos parámetros resultan de analizar y aplicar algún método de estimación o inferencia estadística sobre los datos que se tienen.

En probabilidad y estadística, dado que las poblaciones se caracterizan por medidas descriptivas numéricas llamadas parámetros, el propósito principal es usar la información contenida en una muestra de datos para hacer inferencias acerca de la población de la cual se toma dicha muestra y así, estimar el valor de uno o más parámetros relevantes que servirán en el

modelo matemático planteado para realizar predicciones, tomar medidas de control o estudiar más a fondo el tema en cuestión en una población específica. Muchos de estos métodos de estimación se basan en las distribuciones de probabilidad y para los cálculos debe hacerse uso de un *estimador* del parámetro que se quiere calcular, el cual se define en [16] como *una regla, a menudo expresada como una fórmula, que indica cómo calcular el valor de una estimación con base en las mediciones contenidas en una muestra*.

Algunos métodos de estimación de parámetros son: el método de Mínimos Cuadrados, Máxima Verosimilitud o Estimación por cuantiles. Para la aplicación de estos métodos de estimación, se hace uso de algún lenguaje de programación como Python, Mathematica o Matlab, los cuales permiten realizar una gran cantidad cálculos de una amplia variedad de datos en un corto periodo de tiempo.

1.3.1. Método de mínimos cuadrados (MC)

Uno de los métodos de estimación de parámetros más conocidos es el de Mínimos Cuadrados, en el cual se busca ajustar una curva a cierta cantidad de datos observados. Este método tiene su forma lineal y no lineal, y puede encontrarse en un gran número de libros sobre optimización o métodos numéricos. La estructura de dicho método que presentaremos en esta sección puede encontrarse con más detalle en [12].

En caso de querer ajustar los datos mediante una línea recta, el método es conocido como Mínimos Cuadrados Lineales o Regresión Lineal. Aquí, para un conjunto de $n + 1$ puntos de la forma (x_i, y_i) con $i = 0, 1, 2, \dots, n$, se busca encontrar una función lineal $\hat{y} = mx + b$ de tal forma que se minimice la suma de los errores cuadráticos entre los datos de y_i y los valores $\hat{y}_i$ de la función lineal, esto es minimizar la función descrita en (1.1).

$$\varphi(m, b) = \sum_{i=0}^n (y_i - \hat{y}_i)^2 = \sum_{i=0}^n (y_i - mx_i - b)^2. \quad (1.1)$$

En este caso, los parámetros que se quieren estimar son denominados como la pendiente (m) e intercepto con el eje vertical (b) de la recta de regresión lineal $\hat{y}$. La minimización se realiza calculando las respectivas derivadas parciales φ_m y φ_b e igualándolas a cero. Luego, resolviendo el sistema de ecuaciones resultante aplicando la regla de Cramer se obtiene que

$$m = \frac{1}{d} \text{Det} \begin{bmatrix} r & p \\ q & n+1 \end{bmatrix}; \quad b = \frac{1}{d} \text{Det} \begin{bmatrix} s & r \\ p & q \end{bmatrix}$$

donde

$$p = \sum_{i=0}^n x_i; \quad q = \sum_{i=0}^n y_i; \quad r = \sum_{i=0}^n x_i y_i; \quad s = \sum_{i=0}^n x_i^2.$$

Por otra parte, encontramos el método de Mínimos Cuadrados No Lineales, en el cual, para el mismo conjunto de puntos, se busca encontrar una función $\hat{f}$ de tal manera que la suma de los errores cuadráticos entre los valores y_i y $\hat{f}(x_i)$ sea mínima, es decir, este método busca minimizar la función descrita en (1.2), donde $c_0, c_1, \dots, c_n$ son los coeficientes de la función $\hat{f}$ y por tanto, los parámetros que se quieren estimar.

$$\varphi(c_0, c_1, \dots, c_n) = \sum_{i=0}^m (y_i - \hat{f}(x_i))^2. \quad (1.2)$$

En este caso la función $\hat{f}(x)$ no es necesariamente lineal, un ejemplo de esta puede ser el siguiente:

$$\hat{f}(x) = c_0 \sin(x) + c_1 \cos(x),$$

la cual es lineal solo en sus coeficientes. El tipo de función que pueda ser $\hat{f}$ y la cantidad de coeficientes que tenga determina las restricciones que se deben aplicar al método para poder realizar alguna estimación. En el capítulo 3, se presentan los algoritmos para el algoritmo genético y el empleo de la rutina *lsqcurvefit* de Matlab que se utilizan para la programación del método de estimación de parámetros de mínimos cuadrados no lineales, donde se ajusta cierta cantidad de datos a una función que satisface el modelo en ecuaciones diferenciales ordinarias presentado en el siguiente capítulo.

1.3.2. Algoritmos genéticos (AG)

Como se presenta en [9], los algoritmos genéticos se basan en la mecánica de la selección y genética natural, combinando la supervivencia del más apto y estructuras de cadena con información aleatorizada para formar un algoritmo de búsqueda.

El objetivo principal de un algoritmo genético es encontrar al individuo con las características más adaptables al medio estudiado, que pueda nacer en una población a lo largo del tiempo bajo ciertas restricciones, dichas características deben permitirle tener ventaja de supervivencia sobre el resto de la población, para ello se definen los siguientes conceptos:

- i) **Tamaño de la población:** aquí se fija la cantidad de individuos de la muestra en donde se llevará a cabo el método. Esta cantidad debe ser un valor impar, para así mismo poder escoger en cada generación el más apto, fijarlo y que sea posible realizar una división exacta a la mitad de la población restante.
- ii) **Genes:** dado que este método trabaja combinando estructuras de cadena, cada individuo tiene la misma cantidad de datos, a cada uno de estos datos le llamaremos un genoma. En el momento de la práctica, a cada genoma le corresponde un parámetro específico.

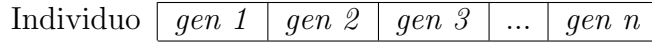


Figura 1.2: Genes como estructura de cadena.

- iii) **Fitness:** Para la selección del más apto, se establece una función restricción a la cual los genes del superviviente se adaptan mejor que los genes del resto de la población en cada generación. En el caso de este trabajo se emplea la función de errores que se quiere minimizar en el método de mínimos cuadrados no lineales.
- iv) **Reproducción:** en esta fase, la población se divide, teniendo en cuenta solo el superviviente y, del resto de la población, la mitad de los que mejor se adaptan a las restricciones establecidas para luego replicarlos y sacar así del sistema a los que menos se adaptan.
- v) **Crossover:** en esta fase, los individuos se organizan por pares (diremos padres) exceptuando el superviviente, cada par dará paso a dos nuevos pobladores en la siguiente generación que tendrán los genes de sus padres mezclados.

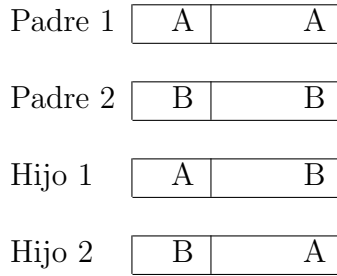


Figura 1.3: Ejemplo de crossover.

- vi) **Mutación:** cada individuo en la nueva generación tiene una probabilidad de que alguno o varios de sus genes muten, de esto se encarga la fase de mutación.

Dadas estas definiciones, es posible notar que hay más de una manera de programar un algoritmo genético, pues el tamaño de la población o la cantidad de genes puede variar, al igual que las restricciones a las que se someten los datos, es decir, el fitness, o la manera en que se realiza el crossover y la mutación. En la sección 3.2, se presenta el algoritmo realizado para este trabajo de grado.

1.4. Datos para el modelo

Puesto que el objetivo de este trabajo de grado es realizar una estimación de parámetros que permita estudiar la dinámica de TB en Cali dada en [5], y comparar los resultados por diferentes métodos empleados, los datos a usar deben ser de la misma localidad. Teniendo en cuenta que la tuberculosis es una enfermedad que puede tener un periodo de latencia extenso, sería infructuoso realizar la estimación de parámetros con datos de un solo año, por ello se hace uso de la información recolectada en un periodo de 5 años desde el 2015 hasta el 2019.

Estos datos fueron reportados por la Secretaría Municipal de Salud Pública de Santiago de Cali y ordenados de manera mensual.

Aunque entre los datos obtenidos se tienen considerados la cantidad acumulada de pacientes en tratamiento, el número de abandonos del tratamiento por mes o la cantidad de muertes por TB, se opta por trabajar con los datos de personas en tratamiento ya que estos tienen una mayor fiabilidad. Si se llegara a considerar cualquiera de los otros conjuntos de datos, como el número de abandonos del tratamiento o las muertes por enfermedad, hay factores externos que pueden haber afectado estos datos, por ejemplo si una persona ingresa a tratamiento y no vuelve después de mucho tiempo puede considerarse un abandono o muerte por enfermedad, pero esta pudo haber muerto por causas ajenas a la TB. Este tipo de casos pueden afectar el estudio de la dinámica notablemente.

En el siguiente capítulo, explicaremos el modelo de la dinámica de la tuberculosis, al cual se le quieren estimar los parámetros, empezando por la definición de las variables, el diagrama de la dinámica de la enfermedad y el modelo en ecuaciones diferenciales ordinarias correspondiente. Se presentará una breve descripción del número básico de reproducción para así realizar el análisis de sensibilidad y la escogencia de los parámetros a estimar.

Capítulo 2

Modelo de transmisión de la Tuberculosis (TB)

2.1. Descripción del modelo

Para efectos de este proyecto, estudiamos el modelo planteado en [5], donde se analiza la enfermedad de la Tuberculosis a través de un modelo compartimental que divide la población en cuatro grupos, dos de estos se explicaron en la sección 1.2, estos son (S) susceptibles e (I) infecciosos, los otros dos compartimientos considerados en el modelo son:

- **Latentes (L):** a esta categoría pertenecen las personas infectadas que no han desarrollado la enfermedad como TB activa.
- **Tratados (T):** en este grupo, se encuentran las personas que reciben alguno de los tratamientos, para controlar los bacilos de TB en su cuerpo.

Cada uno de estos grupos se denotan con las funciones $S(t)$, $L(t)$, $I(t)$ y $T(t)$ respectivamente para un tiempo $t \geq 0$, por lo que la población total en un determinado tiempo t es

$$N(t) = S(t) + L(t) + I(t) + T(t). \quad (2.1)$$

En la Figura 2.1, se visualiza el diagrama que representa la dinámica de la transmisión de la enfermedad, el cual muestra las transiciones que hay entre cada uno de los cuatro grupos en que se dividió la población. Esta información se sintetiza en el sistema de ecuaciones diferenciales (2.2).

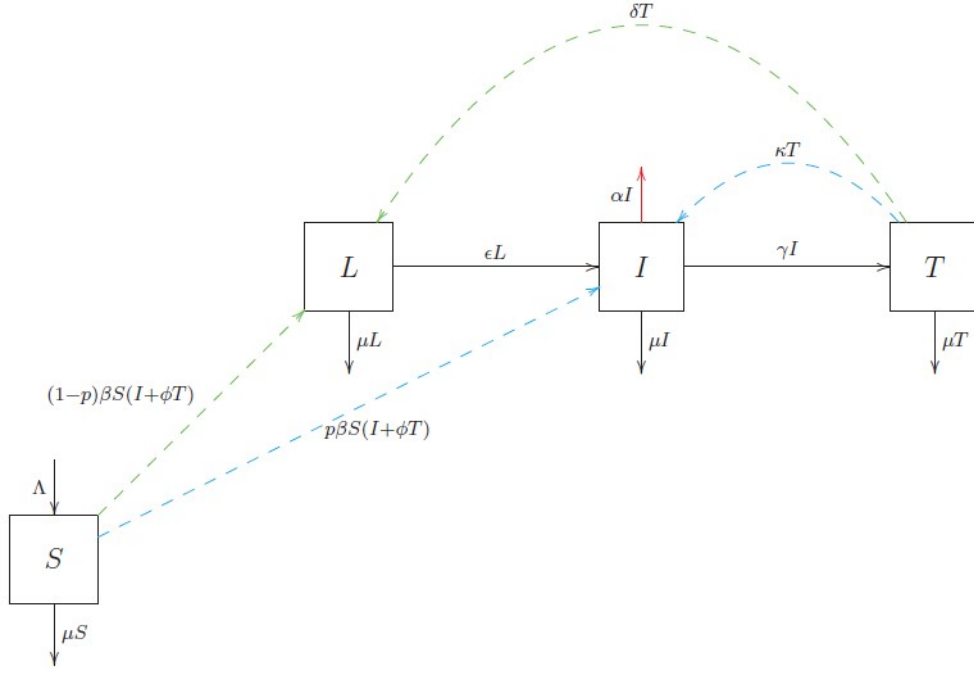


Figura 2.1: Diagrama del modelo de la dinámica de la tuberculosis dada en [5].

$$\left\{ \begin{array}{l} \frac{dS}{dt} = \Lambda - \beta(I + \phi T)S - \mu S \\ \frac{dL}{dt} = (1-p)\beta(I + \phi T)S + \delta T - \epsilon L - \mu L \\ \frac{dI}{dt} = p\beta(I + \phi T)S + \epsilon L + \kappa T - \gamma I - \alpha I - \mu I \\ \frac{dT}{dt} = \gamma I - \delta T - \kappa T - \mu T \end{array} \right. \quad (2.2)$$

Para el sistema de ecuaciones diferenciales (2.2), se tienen las siguientes condiciones iniciales $S(0) = S_0$, $L(0) = L_0$, $I(0) = I_0$, $T(0) = T_0$ y los parámetros están descritos en la Tabla 2.1.

Con el fin de establecer una cota superior para la población total, al derivar la ecuación (2.1) se obtiene que $\frac{dN}{dt}$ es la suma de las cuatro ecuaciones del sistema, es decir,

$$\frac{dN}{dt} = \Lambda - \alpha I - \mu N. \quad (2.3)$$

De (2.3), se observa que $\frac{dN}{dt} \leq \Lambda - \mu N$ y por tanto una aproximación al tamaño total de la población es $N(t) \approx \frac{\Lambda}{\mu}$, este resultado se obtiene al aplicar el método de variables separables

Parámetro	Descripción
Λ	Reclutamiento de individuos humanos
μ	Tasa de mortalidad natural
β	Tasa de transmisión de TB
ϕ	Reducción de infecciosidad por tratamiento
p	Fracción de infecciones rápidas
ϵ	Tasa de desarrollo lento de la infección
δ	Tasa de tratamiento completo
γ	Tasa de reclutamiento para tratamiento
α	Tasa de mortalidad por la enfermedad
κ	Tasa de abandono de tratamiento

Tabla 2.1: Descripción de los parámetros del modelo de TB dado en [5].

y tendiendo t a una cantidad de tiempo muy grande.

De esta manera, se pueden interpretar cada una de las ecuaciones del modelo como sigue:

- En la primera, la variación de la población susceptible, el parámetro Λ hace referencia a la tasa de personas que ingresan al modelo, esto suponiendo que la enfermedad no se transmite de manera vertical; es decir, que una embarazada infectada con TB activa no transmite la enfermedad al bebé que está por nacer. De esta misma, se resta la tasa de personas susceptibles que hayan tenido contacto con los infecciosos, βIS , o con las personas que aún están en tratamiento, $\beta \phi TS$ y, también se resta la tasa de personas susceptibles que mueren por causa natural, μS .
- En la segunda, la variación de la población latente, se encuentra la fracción de personas anteriormente susceptibles que al tener contacto con algún individuo infeccioso o en tratamiento, pasan por un periodo de latencia teniendo un desarrollo lento de la enfermedad, $(1 - p)\beta(I + \phi T)S$; aquí también se suma el término referente a la cantidad de personas que finalizaron el tratamiento, δT , y se restan las tasas referentes a quienes terminan el largo periodo de latencia y desarrollan la enfermedad como TB activa, ϵL , y la de mortalidad por causa natural μL .
- En la tercera, la variación de la población infecciosa, se suma la fracción de personas que, al contrario de la anterior, luego del contacto con individuos infecciosos y tratados, desarrollan la enfermedad de manera rápida; es decir, que tuvieron un periodo de latencia muy corto, $p\beta(I + \phi T)S$; también la referente a quienes después de un largo periodo de latencia desarrollan la enfermedad, ϵL , y la de quienes abandonan el tratamiento, κT . De aquí se restan las tasas de personas reclutadas para tratamiento, la de muertos por la enfermedad y la de muertos de manera natural, γI , αI y μI , respectivamente.
- En la cuarta ecuación, la variación de la población en tratamiento, se suma la tasa de quienes fueron reclutados para tratamiento, γI , y se restan las tasas referentes a los abandonos de tratamiento, finalización de tratamiento y la de muerte por causa natural.

2.2. Número básico de reproducción ($\mathcal{R}_0$)

En epidemiología, una de las principales medidas relacionadas con la velocidad de propagación de una enfermedad en la población es el llamado número básico de reproducción $\mathcal{R}_0 > 0$. Esta cantidad expresa el número promedio esperado de infecciones secundarias causadas por la introducción de un individuo infectado en una población de susceptibles en su periodo de infecciosidad [13].

A partir del punto de equilibrio libre de enfermedad, $\mathbf{E}^0 = \left(\frac{\Lambda}{\mu}, 0, 0, 0\right)$, es posible calcular el $\mathcal{R}_0$ como el radio espectral de la matriz de la próxima generación [15]. El procedimiento a seguir para el cálculo de los puntos de equilibrio y del número básico de reproducción se encuentra en el análisis cualitativo del sistema desarrollado en [5], donde se obtiene que

$$\mathcal{R}_0 = \frac{\beta \frac{\Lambda}{\mu} ((\delta + \kappa + \mu) + \phi\gamma)}{(\epsilon + \mu)[(\delta + \kappa + \mu)(\alpha + \mu) + \gamma\mu] + \gamma\delta\mu} [p(\epsilon + \mu) + (1 - p)\epsilon]. \quad (2.4)$$

Con el resultado obtenido en (2.4), es posible hacer deducciones respecto al comportamiento de la enfermedad según las variaciones de los parámetros, para ello es necesario saber cuáles de éstos parámetros causan grandes cambios en el resultado del $\mathcal{R}_0$ con pequeñas perturbaciones dentro de su rango, esto se ve en el análisis de sensibilidad y nos permite identificar para cuáles de estos parámetros es más sensible el modelo.

2.3. Sensibilidad del $\mathcal{R}_0$

El cálculo del índice de sensibilidad del número básico de reproducción se realiza mediante la definición presentada a continuación. Esta puede encontrarse con mayores detalles en [6].

Definición 2.3.1. El índice de sensibilidad normalizado de una variable u , que depende diferenciablemente de un parámetro p , se define como:

$$\Upsilon_p^u := \frac{\partial u}{\partial p} \times \frac{p}{u}. \quad (2.5)$$

Para realizar el análisis de sensibilidad del número básico de reproducción $\mathcal{R}_0$, se debe establecer un valor para cada parámetro y hacer los cálculos respectivos alrededor de este; en este documento, dicho valor se presenta en la segunda columna de la Tabla 2.2. Puesto que la tasa de abandono de tratamiento (κ) se presenta como una cantidad variable dentro de un rango determinado, a diferencia de los demás parámetros que tienen un punto de partida o condición inicial, se realizan las simulaciones para diferentes valores dentro de su rango, los cuales por escogencia arbitraria son 0,025; 0,0083; 0,0025 y 0.000083.

	Parámetro	Valor	Cota Inferior	Cota Superior
1	α	0,00025	0,0019	0,025
2	β	$0,14 \times 10^{-7}$	$\frac{0,01}{32000(12)/(\frac{1}{77})}$	$\frac{1}{32000(12)/(\frac{1}{77})}$
3	γ	$\frac{1}{6}$	0,025	0,2
4	δ	0,125	0,094	0,125
5	ϵ	0,004	0,0003	0,02
6	μ	$\frac{1}{77(12)}$	$\frac{1}{80(12)}$	$\frac{1}{74(12)}$
7	Λ	32000	30000	34000
8	ϕ	0,02	0	$\frac{1}{12}$
9	p	0,05	0,025	0,3
10	κ	Varía	0	0,04

Tabla 2.2: Rangos para los parámetros del modelo presentados en [5] ajustados a los datos de Cali en meses.

Siguiendo entonces la definición, se calculan las derivadas parciales del $\mathcal{R}_0$ respecto a cada parámetro y se emplea la expresión (2.5) con la ayuda de la herramienta matemática **Wolfram Mathematica**. El índice de sensibilidad de cada parámetro para las variaciones de κ se registra en la Tabla 2.3.

κ	α	β	γ	δ	ϵ	μ	Λ	ϕ	p	κ
0,025	-0,08	1	-0,87	-0,14	0,87	-1,94	1	0,02	0,01	0,14
0,0083	-0,07	1	-0,88	-0,05	0,88	-1,94	1	0,03	0,01	0,05
0,0025	-0,07	1	-0,88	-0,01	0,88	-1,94	1	0,03	0,01	0,02
0,000083	-0,06	1	-0,88	0,002	0,89	-1,94	1	0,03	0,01	0,001

Tabla 2.3: Índice de sensibilidad del $\mathcal{R}_0$ respecto a cada parámetro.

De la Tabla 2.3, se deduce que los parámetros que tienen una influencia mayor en los cambios del $\mathcal{R}_0$ con variaciones pequeñas de ellos mismos son β , γ , ϵ , μ y Λ . Seguidamente, otro de los parámetros que tiene una influencia grande pero debajo de la media en la sensibilidad del $\mathcal{R}_0$ es la tasa de mortalidad por la enfermedad α .

Para efectos de este trabajo de grado, se realiza la estimación de los parámetros que se presentan en la Tabla 2.4, pues los demás pueden ser tomados desde la literatura como lo son la tasa de mortalidad natural μ que se calcula a partir de la esperanza de vida de una persona en Cali, o el número de individuos humanos reclutados para el modelo, Λ , ya que son cantidades que no dependen directamente de la enfermedad y difícilmente se pueden deducir a partir de los datos. Estos valores fijos para cada parámetro son los mismos con que se desarrolla el cálculo del índice de sensibilidad.

El parámetro ϕ , por otra parte, aunque no presenta un índice de sensibilidad alto en los cálculos realizados, compromete información sobre las personas en tratamiento ya que es

la tasa de reducción de infecciosidad de la enfermedad, de allí que veamos la necesidad de incluirlo entre los parámetros que se quieren estimar.

Parámetro	Descripción
β	Tasa de transmisión de TB
ϕ	Reducción de infecciosidad por tratamiento
ϵ	Tasa de desarrollo lento de la infección
γ	Tasa de reclutamiento para tratamiento
α	Tasa de mortalidad por la enfermedad

Tabla 2.4: Conjunto de parámetros a estimar.

2.4. Condiciones Iniciales

Según lo reportado por el Ministerio de Salud en el Boletín Epidemiológico Semanal [11] para el año 2019, la incidencia de tuberculosis en Colombia se encontraba en 27,69 por 100.000 habitantes. Para este mismo año, siguiendo los reportes del Departamento Administrativo Nacional de Estadística (DANE), la población total de Cali fue de 2.241.491 personas. Adaptando entonces estos datos al modelo, podemos definir las condiciones iniciales S_0 , L_0 , I_0 y T_0 como se muestra en la siguiente tabla:

Variable	Valor Inicial
I_0	621
T_0	342
L_0	$\frac{I_0}{0,05}$
S_0	$(2,241 \times 10^6) - L_0 - I_0 - T_0$

Tabla 2.5: Condiciones iniciales del sistema.

2.5. Modelo de TB normalizado

A partir del modelo descrito en la sección 2.1 y teniendo en cuenta la aproximación dada para la población total $N(t) \approx \frac{\Lambda}{\mu}$, podemos realizar su respectiva normalización. Sea $N^*(t) = \frac{\Lambda}{\mu}$ la población total en un tiempo t , determinada cuando la tasa de mortalidad por la enfermedad es nula, es decir $\alpha = 0$ y sean s , l , i y x las variables del sistema normalizado definidas de la forma

$$s = \frac{S}{N^*} = \frac{\mu S}{\Lambda}; \quad l = \frac{L}{N^*} = \frac{\mu L}{\Lambda}; \quad i = \frac{I}{N^*} = \frac{\mu I}{\Lambda}; \quad x = \frac{T}{N^*} = \frac{\mu T}{\Lambda}$$

y por tanto, sus derivadas respecto al tiempo t son

$$\frac{ds}{dt} = \frac{\mu}{\Lambda} \frac{dS}{dt}; \quad l = \frac{\mu}{\Lambda} \frac{dL}{dt}; \quad i = \frac{\mu}{\Lambda} \frac{dI}{dt}; \quad x = \frac{\mu}{\Lambda} \frac{dT}{dt}.$$

Con estos resultados unidos al sistema de ecuaciones (2.2), se obtiene el nuevo sistema de ecuaciones diferenciales normalizado (2.6), donde $\eta = \beta N^*$ es un nuevo parámetro que, al analizarse la expresión de β dada en [5], se define como

$\eta := (\# \text{ de contactos por persona por unidad de tiempo}) \times (\text{probabilidad de contagio}).$

$$\left\{ \begin{array}{l} \frac{ds}{dt} = \mu - \eta(i + \phi x)s - \mu S \\ \frac{dl}{dt} = (1 - p)\eta(i + \phi x)s + \delta x - \epsilon l - \mu l \\ \frac{di}{dt} = p\eta(i + \phi x)s + \epsilon l + \kappa x - \gamma i - \mu i \\ \frac{dx}{dt} = \gamma i - \delta x - \kappa x - \mu x \end{array} \right. \quad (2.6)$$

Capítulo 3

Programas y Resultados

En este capítulo, se presentan los algoritmos utilizados para la programación del método de mínimos cuadrados no lineales por medio de la rutina de *lsqcurvefit* y el algoritmo genético en el software de programación de **MATLAB** junto con los resultados obtenidos para los parámetros a estimar y el error arrojado por cada método empleado. Los programas se ejecutan para cada distinto valor de κ elegidos en la sección 2.3, pero presentamos los resultados para $\kappa = 0,000083$, pues entre los valores seleccionados fue el único que permitió encontrar un mínimo local en el programa de mínimos cuadrados no lineales.

Dado que para el modelo normalizado se asume $\alpha = 0$, los parámetros a estimar de la Tabla 2.4 se reducen a β , γ , ϵ y ϕ . Sin embargo, para realizar la estimación de la tasa de transmisión de la enfermedad, β , a partir del modelo normalizado, debe calcularse inicialmente un valor para η y seguidamente aplicar la expresión $\beta = \frac{\eta}{N^*}$. Por otro lado, en las primera, segunda y tercera ecuaciones del modelo aparece el término cruzado $\eta\phi$ donde ambos son parámetros distintos que queremos estimar. Puesto que η es independiente del valor de ϕ y además se encuentra en términos no colineales entre todo el sistema de ecuaciones, nos permite realizar la estimación de $\eta\phi$ asumiéndolo como un nuevo parámetro y, dividiendo este último resultado entre el valor estimado para η conseguimos el resultado del parámetro ϕ . Los rangos para el nuevo parámetro y el término cruzado, tomados para la programación y ejecución de los programas se obtienen a partir de la Tabla 2.2 y están consignados en la Tabla 3.1.

Parámetro	Cota Inferior	Cota Superior
η	0,00083	0,083
$\eta\phi$	0	0,0069

Tabla 3.1: Rangos de los parámetros a estimar en el modelo normalizado.

3.1. Rutina de *lsqcurvefit* de Matlab

Para la programación del método de mínimos cuadrados no lineales que se desarrolla en este trabajo de grado usando las rutinas *lsqcurvefit* y *ode45* de Matlab, se realizan los siguientes pasos:

- **Dinámica:** aquí se define el sistema de ecuaciones diferenciales ordinarias normalizado, sus variables y cada uno de sus parámetros, estableciendo como un vector variable que tiene como componentes los parámetros a estimar y dejando los demás parámetros como cantidades globales.
- **Función a ajustar:** introducimos la ecuación de tratados acumulados como la solución del problema de valores iniciales determinado por la ecuación (3.1) y $x^*(0) = 0$. Como resultado, el sistema (2.6) aumenta en una quinta ecuación diferencial ordinaria adicional.

$$\frac{dx^*}{dt} = \gamma i. \quad (3.1)$$

- **Lectura de datos:** se ingresan al programa los datos de personas en tratamiento explicados en la sección 1.4, desde el archivo de excel con el comando *xlsread*.
- **Condiciones iniciales:** en este paso, se establecen las condiciones iniciales para el sistema de ecuaciones diferenciales normalizado 2.6, estas se presentan en la Tabla 3.2 y son adaptadas desde los datos de la Tabla 2.5.

Variable	Valor Inicial
i_0	0,00025
x_0	0,00014
l_0	0,00500
s_0	0,99461

Tabla 3.2: Condiciones iniciales del sistema normalizado.

- **Rangos:** aquí se codifican los rangos de cada parámetro a estimar y se fija el valor de los parámetros tomados de la literatura. Estos rangos y valores fijos se encuentran en las Tablas 2.2 y 3.1. También seleccionamos el punto de partida para cada parámetro, estos se muestran en la Tabla 3.3.
- **Ajuste:** se resuelve la ecuación (3.1) con los valores iniciales de los parámetros. Luego, utilizando los datos de personas en tratamiento acumuladas y la función de tratados acumulados, se realiza el ajuste con el comando *lsqcurvefit*, el cual se encarga de ajustar una función a un conjunto de datos mediante el algoritmo de *Levenberg-Marquard* aplicando el método de mínimos cuadrados no lineales.
- **Solución de la dinámica:** haciendo uso del comando *ode45* se resuelve la dinámica para el conjunto de parámetros calculados luego del ajuste.

- **Cálculo de errores:** comparamos cada uno de los puntos de los datos obtenidos para personas en tratamiento y en tratamiento acumuladas, con las funciones resultantes para tratados y tratados acumulados respectivamente, midiendo las distancias al cuadrado entre cada dato y su función de ajuste correspondiente. Luego de esto, presentamos los resultados gráficamente y el valor de la suma de los errores cuadráticos para los datos de tratados acumulados.
- **Gráficas y resultados:** se muestran los resultados de los parámetros estimados por el método. También se presentan la gráfica de dispersión de los datos trabajados junto con sus respectivos ajustes, errores y soluciones del sistema de ecuaciones diferenciales. Igualmente se muestra el error cuadrático resultante del ajuste realizado.

3.2. Algoritmo genético

El algoritmo genético, presentado en la sección 1.3.2, se centra en encontrar el mejor individuo que puede nacer de una población con el paso del tiempo utilizando una función que los clasifique dentro de la población según su adaptabilidad a las restricciones del medio, por lo que para la programación de este algoritmo es necesario definir algunas constantes nuevas y adaptarlo para la estimación de los parámetros elegidos:

- **TPOB:** indica el tamaño de la población, es decir cuántos individuos se quieren en la muestra aleatoria.
- **Parámetros:** esta cantidad hace referencia al número de genes o características de cada individuo. Para efectos de este trabajo de grado se toman como la cantidad de parámetros que se quieren estimar.

Habiendo definido tales cantidades se hace uso del comando *lhsdesign* de MATLAB para generar el hipercubo (matriz) que el algoritmo requiere, con TPOB y Parametros como su cantidad de filas y columnas respectivamente. Este código asigna valores aleatorios en el rango (0,1) para cada posición en la matriz, por lo que es necesario establecer una transformación que permita, para cada columna, tomar valores en el rango designado de cada parámetro; estos rangos se encuentran en las Tablas 2.2 y 3.1.

Sean entonces b y a las cotas superior e inferior del parámetro en la columna i , y por tanto $b - a$ la amplitud de este rango. Si x es un dato en el hipercubo de la columna i , entonces el nuevo valor asignado será $y = (b - a)x + a$. Llamaremos *muestra* a la matriz del hipercubo modificada con los rangos de cada parámetro.

Ahora teniendo la muestra se puede proceder con el resto del código:

- **Lectura de datos:** se ingresan al programa los datos de personas en tratamiento explicados en la sección 1.4.
- **Condiciones iniciales:** se establecen condiciones iniciales para el sistema de ecuaciones 2.6 como en la Tabla 3.2.

- **Rangos:** aquí se codifican los rangos de cada parámetro a estimar y se fija el resto tomando los datos de la Tabla 2.2.
- **Bucle:** en esta parte del programa se debe establecer un criterio de parada, ya que una sola repetición del bucle corresponde a una sola generación del algoritmo genético y, como resultado, los datos encontrados pueden no cumplir con todas las restricciones establecidas y consiguientemente no optimizar el problema. Este criterio de parada es necesario dado que el algoritmo puede seguir ejecutándose indefinidamente por la aleatoriedad de algunos datos que aquí se consideran. Dentro de este bucle, se realiza la dinámica, el fitness, el crossover y la mutación.
 - **Dinámica:** para cada fila en la muestra se resuelve el sistema de ecuaciones diferenciales usando el comando *ode45*.
 - **Fitness:** se calcula el error cuadrático entre la solución de la ecuación de tratados y el conjunto de datos de personas en tratamiento cada que se revuelve la dinámica. Se agrega el vector error como una nueva columna a la matriz de la muestra para ordenar todas las filas de menor a mayor error y fijar la primera fila que corresponde al conjunto de parámetros más apto. Esta función (fitness) es justamente la función de errores que se minimiza en el método de mínimos cuadrados no lineales.
 - **Reproducción:** ignorando la primera fila se eligen la primera mitad de las filas restantes en la muestra y se duplican estos valores en su segunda mitad, reemplazando los menos aptos.
 - **Crossover:** ignorando nuevamente la primera fila, se escogen arbitrariamente *padre 1* y *padre 2* de la muestra restante, y se elige un punto de corte, también aleatorio, entre 1 y la cantidad de parámetros a estimar para así combinar los datos (genes) de ambos padres y dar paso a dos nuevos hijos en la nueva muestra (generación).
 - **Mutación:** se fija una probabilidad de mutación para cada uno de los nuevos individuos, en este caso se escoge esa probabilidad con un valor de 0.01 y, para cada fila en la nueva muestra se elige un valor aleatorio entre 0 y 1. Si el valor es menor que la probabilidad de mutación, se elige un valor aleatorio entre 1 y la cantidad de parámetros a estimar y el gen en dicha posición cambia por otro valor aleatorio dentro del rango del parámetro en esa columna.
- **Gráficas:** se presentan los resultados obtenidos después de ejecutar el bucle hasta su criterio de parada, esto es el error del método, las gráficas de dispersión de puntos usados, el valor de los parámetros calculados y las soluciones del sistema de ecuaciones diferenciales.

3.3. Resultados

Para cada uno de los algoritmos empleados, hacemos uso del modelo normalizado presentado en la sección 2.5, de lo contrario, si se emplea el modelo sin normalizar, el programa de mínimos cuadrados resuelto por la rutina de *lsqcurvefit* no encuentra el mínimo local debido a las

diferentes escalas de las variables. También se consideran diferentes condiciones iniciales para los parámetros que se quieren estimar dentro de su rango establecido, para así poder analizar y comparar la convergencia de cada algoritmo, es decir los resultados estimados, aplicando pequeñas perturbaciones en las condiciones iniciales de cada parámetro, permitiendo así realizar conclusiones sobre el mínimo local encontrado y la sensibilidad de los resultados a estas. Estas condiciones iniciales son tomadas en los programas como los vectores presentados en la Tabla 3.3.

Vector de condiciones iniciales	η	γ	ϵ	$\eta\phi$
x_1	0,033880	0,16667	0,0041667	0,0007
x_2	0,033880	0,08081	0,0041667	0,0018
x_3	0,035336	0,08081	0,0041667	0,0018
x_4	0,035336	0,08081	0,0065796	0,0018
x_5	0,035336	0,08081	0,0065796	0,0060
x_6	0,030256	0,08081	0,0065796	0,0060
x_7	0,030256	0,08081	0,0065796	0,0050
x_8	0,030256	0,08989	0,0065796	0,0050
x_9	0,030256	0,08989	0,0064417	0,0050
x_{10}	0,030256	0,08989	0,0064417	0,0040

Tabla 3.3: Vectores de condiciones iniciales para los parámetros.

El programa de cada método empleado, se ejecuta un total de diez veces, cubriendo todas las condiciones iniciales establecidas para los parámetros. Las gráficas aquí presentadas corresponden solamente al vector de condiciones iniciales x_1 , pues visualmente no se perciben cambios usando un vector u otro de las soluciones del sistema y el ajuste, esto debido a la escala de las mismas, mostrando que los resultados no son sensibles a los cambios en las condiciones iniciales.

A continuación, se presentan los resultados obtenidos con el método de mínimos cuadrados, las gráficas de las soluciones del sistema, el ajuste realizado y las gráficas de dispersión de los errores cuadráticos obtenidos. Los resultados numéricos para cada vector de condiciones iniciales utilizado se encuentran en la Tabla 3.4.

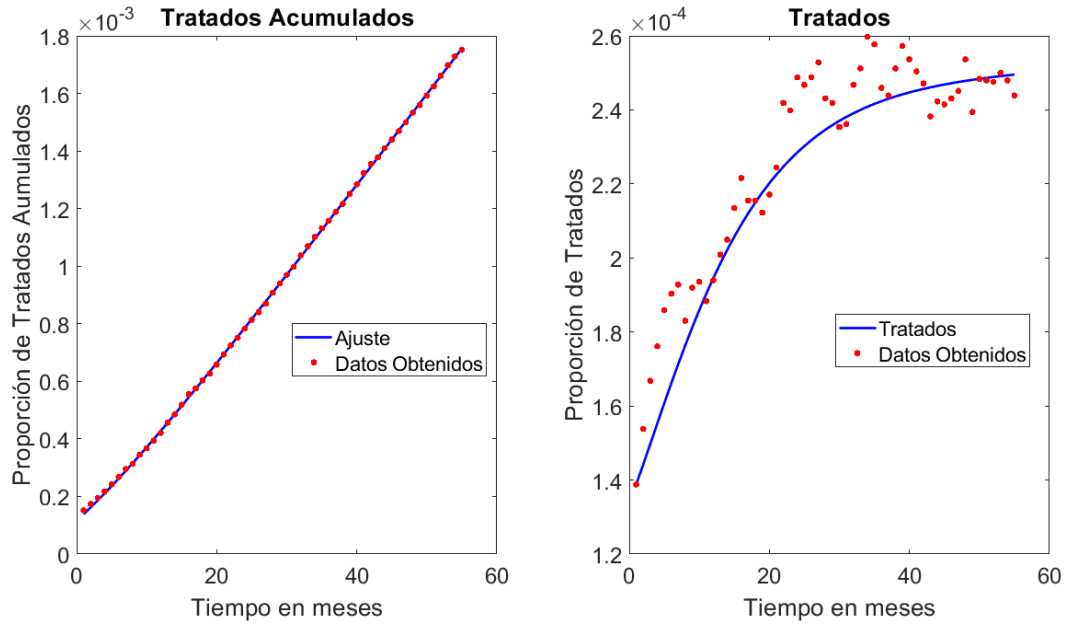


Figura 3.1: Ajuste realizado por *lsqcurvefit* y solución de la ecuación de tratados en el sistema (2.2).

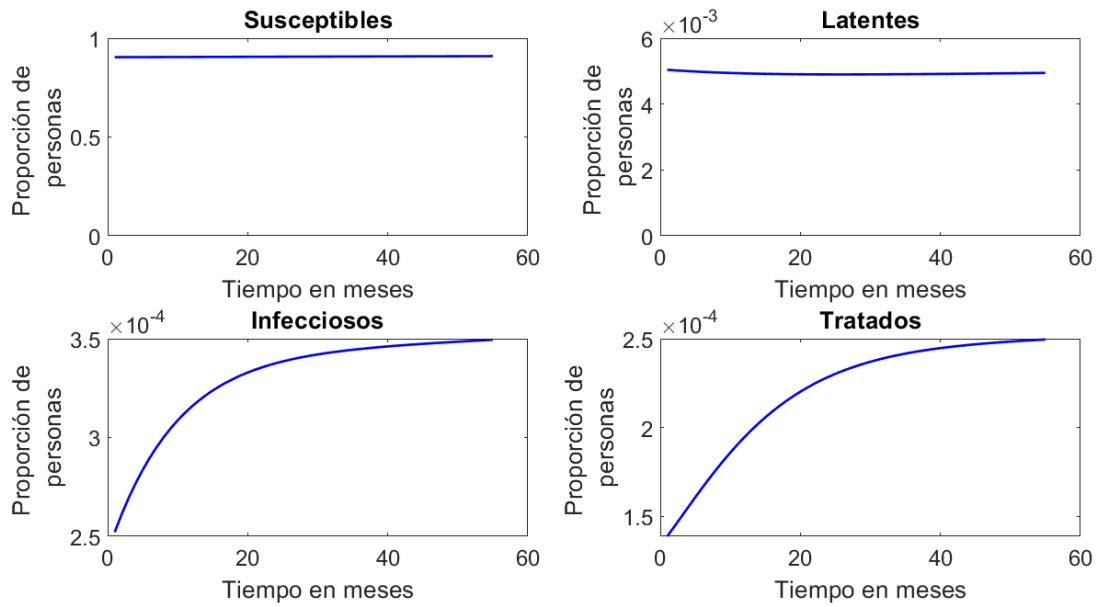


Figura 3.2: Soluciones del sistema (2.2) con los parámetros encontrados por *lsqcurvefit*.

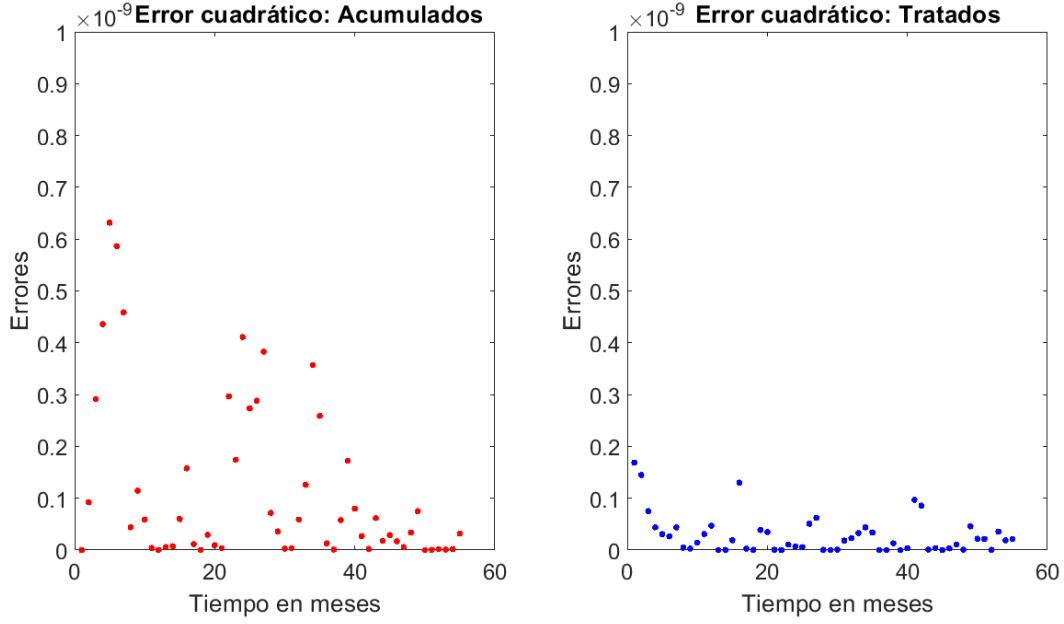


Figura 3.3: Errores cuadráticos de *lsqcurvefit*.

Vector	η	γ	ϵ	ϕ	β	Error Cuadrático
x_1	$2,22 \times 10^{-2}$	$9,07 \times 10^{-2}$	$6,43 \times 10^{-3}$	0,313	$9,01 \times 10^{-9}$	$1,53 \times 10^{-9}$
x_2	$3,38 \times 10^{-2}$	$8,17 \times 10^{-2}$	$6,62 \times 10^{-3}$	0,051	$1,37 \times 10^{-8}$	$9,07 \times 10^{-9}$
x_3	$3,52 \times 10^{-2}$	$8,18 \times 10^{-2}$	$6,61 \times 10^{-3}$	0,049	$1,43 \times 10^{-8}$	$9,38 \times 10^{-9}$
x_4	$3,53 \times 10^{-2}$	$8,09 \times 10^{-2}$	$6,64 \times 10^{-3}$	0,051	$1,43 \times 10^{-8}$	$1,09 \times 10^{-8}$
x_5	$3,53 \times 10^{-2}$	$8,09 \times 10^{-2}$	$6,61 \times 10^{-3}$	0,170	$1,43 \times 10^{-8}$	$1,15 \times 10^{-8}$
x_6	$3,03 \times 10^{-2}$	$8,09 \times 10^{-2}$	$6,66 \times 10^{-3}$	0,198	$1,23 \times 10^{-8}$	$1,02 \times 10^{-8}$
x_7	$3,03 \times 10^{-2}$	$8,09 \times 10^{-2}$	$6,67 \times 10^{-3}$	0,165	$1,23 \times 10^{-8}$	$1,01 \times 10^{-8}$
x_8	$3,02 \times 10^{-2}$	$9,00 \times 10^{-2}$	$6,39 \times 10^{-3}$	0,165	$1,23 \times 10^{-8}$	$1,72 \times 10^{-9}$
x_9	$3,03 \times 10^{-2}$	$8,99 \times 10^{-2}$	$6,39 \times 10^{-3}$	0,165	$1,23 \times 10^{-8}$	$1,75 \times 10^{-9}$
x_{10}	$3,03 \times 10^{-2}$	$8,99 \times 10^{-2}$	$6,39 \times 10^{-3}$	0,132	$1,23 \times 10^{-8}$	$1,72 \times 10^{-9}$

Tabla 3.4: Tabla de resultados para las simulaciones de la rutina *lsqcurvefit* con las condiciones iniciales de los parámetros dadas en la Tabla 3.3.

Por otra parte, para el algoritmo genético, se toma un tamaño de la población TPOB = 5001 y un criterio de parada para el bucle con 50 repeticiones, los resultados al ejecutar el programa con el primer vector de condiciones iniciales para los parámetros son los que se muestran a continuación. Los resultados numéricos son registrados en la Tabla 3.5.

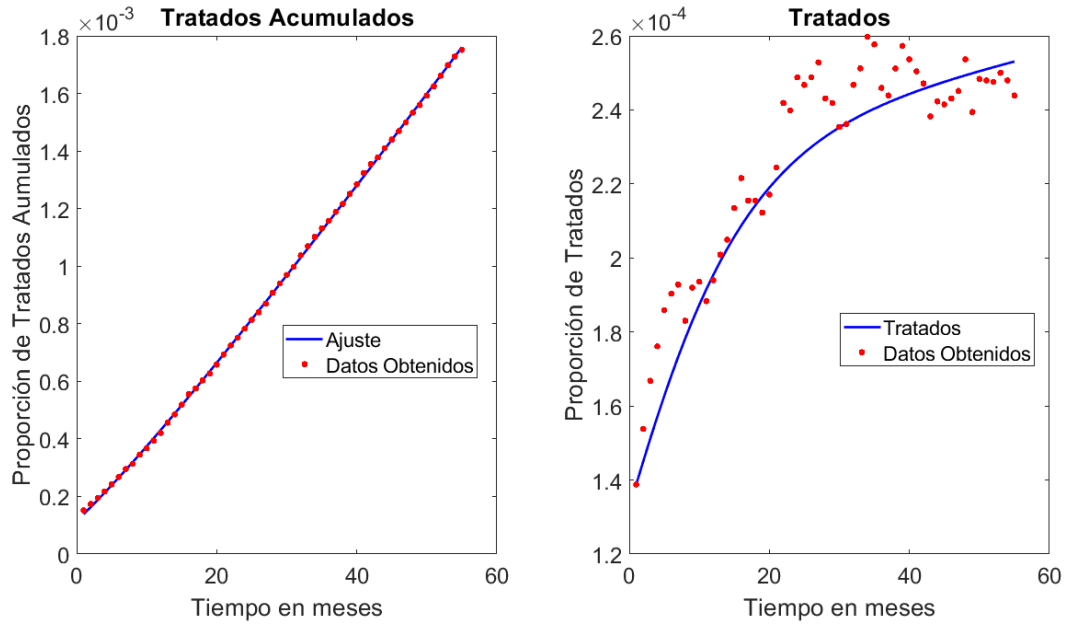


Figura 3.4: Ajuste realizado por el algoritmo genético y solución de la ecuación de tratados en el sistema (2.2).

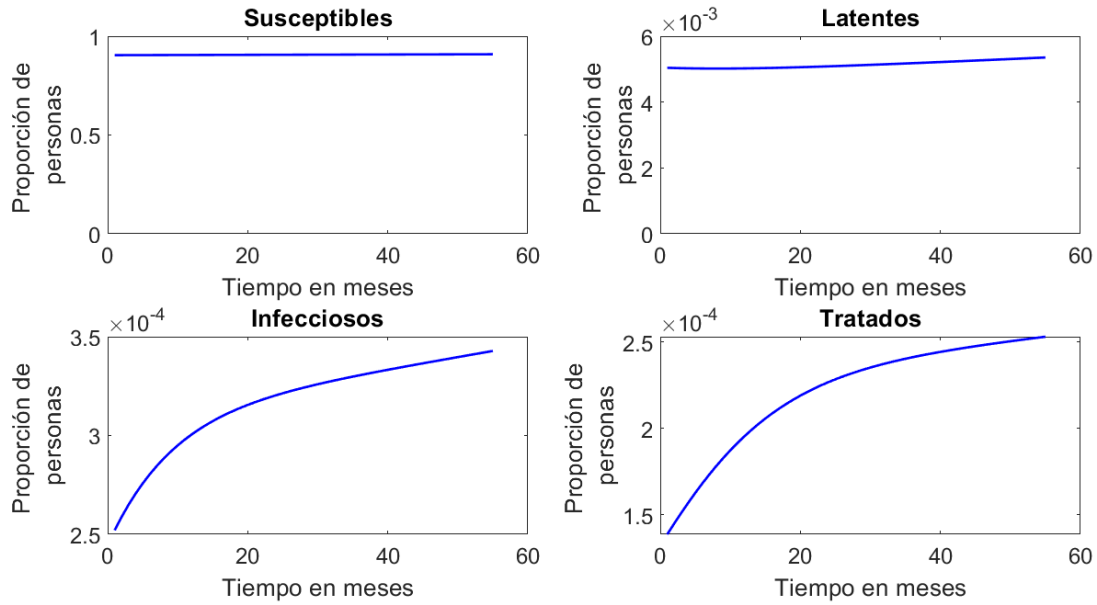


Figura 3.5: Soluciones del sistema (2.2) con los parámetros encontrados por el algoritmo genético.

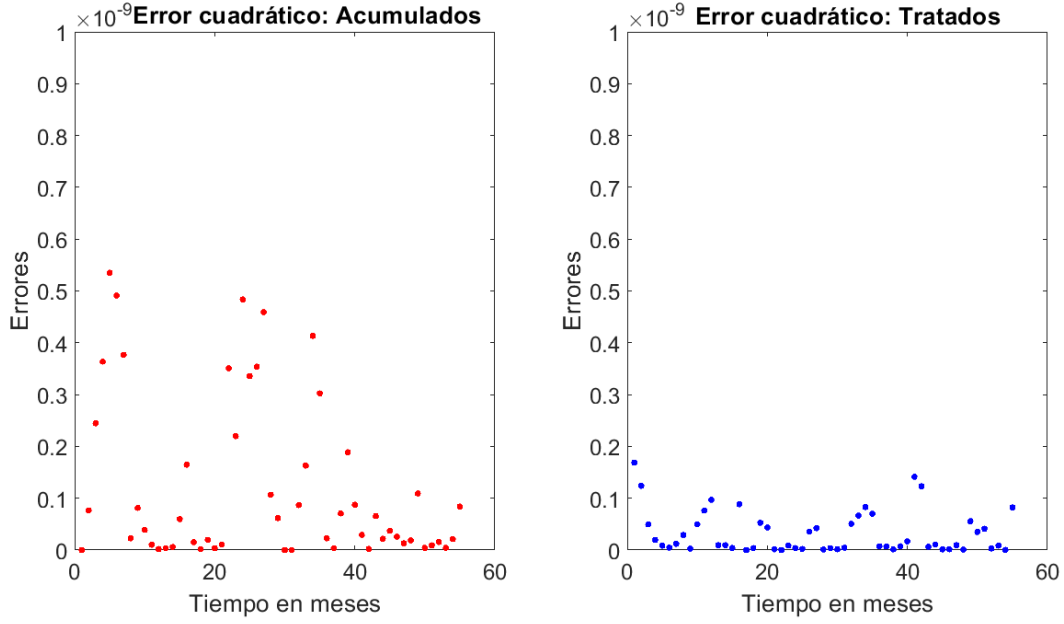


Figura 3.6: Errores cuadráticos del algoritmo genético.

Vector	η	γ	ϵ	ϕ	β	Error Cuadrático
x_1	$2,31 \times 10^{-2}$	$9,02 \times 10^{-2}$	$6,45 \times 10^{-3}$	0,171	$9,36 \times 10^{-9}$	$1,84 \times 10^{-9}$
x_2	$3,52 \times 10^{-2}$	$9,16 \times 10^{-2}$	$6,30 \times 10^{-3}$	0,140	$1,43 \times 10^{-8}$	$1,79 \times 10^{-9}$
x_3	$2,02 \times 10^{-2}$	$9,02 \times 10^{-2}$	$6,47 \times 10^{-3}$	0,282	$8,20 \times 10^{-9}$	$1,56 \times 10^{-9}$
x_4	$2,20 \times 10^{-2}$	$9,07 \times 10^{-2}$	$6,44 \times 10^{-3}$	0,292	$8,94 \times 10^{-9}$	$1,61 \times 10^{-9}$
x_5	$1,84 \times 10^{-2}$	$8,97 \times 10^{-2}$	$6,51 \times 10^{-3}$	0,228	$7,46 \times 10^{-9}$	$1,80 \times 10^{-9}$
x_6	$3,14 \times 10^{-2}$	$9,11 \times 10^{-2}$	$6,36 \times 10^{-3}$	0,079	$1,28 \times 10^{-8}$	$1,77 \times 10^{-9}$
x_7	$1,90 \times 10^{-2}$	$8,99 \times 10^{-2}$	$6,53 \times 10^{-3}$	0,003	$7,71 \times 10^{-9}$	$1,81 \times 10^{-9}$
x_8	$2,24 \times 10^{-2}$	$9,07 \times 10^{-2}$	$6,44 \times 10^{-3}$	0,215	$9,07 \times 10^{-9}$	$1,68 \times 10^{-9}$
x_9	$2,64 \times 10^{-2}$	$9,07 \times 10^{-2}$	$6,42 \times 10^{-3}$	0,094	$1,07 \times 10^{-8}$	$1,59 \times 10^{-9}$
x_{10}	$3,10 \times 10^{-2}$	$9,13 \times 10^{-2}$	$6,37 \times 10^{-3}$	0,027	$1,26 \times 10^{-8}$	$1,84 \times 10^{-9}$

Tabla 3.5: Tabla de resultados para las simulaciones del algoritmo genético (AG) con las condiciones iniciales de los parámetros dadas en la Tabla 3.3.

3.4. Discusión de resultados

El método de mínimos cuadrados, aplicando la rutina de *lsqcurvefit* tarda un promedio de 2,5 segundos en ejecutarse y cada una de las simulaciones realizadas para los vectores de condiciones iniciales de la Tabla 3.3, presenta un error cuadrático en el rango de $1,53 \times 10^{-9}$ hasta $1,15 \times 10^{-8}$, siendo x_1 , x_8 y x_{10} las tres condiciones iniciales para los parámetros que permiten realizar el ajuste con error más bajo. Por otro lado, el algoritmo genético tarda en promedio 10,5 minutos en ejecutarse con rangos de error cuadrático para cada simulación

entre $1,56 \times 10^{-9}$ y $1,84 \times 10^{-9}$, siendo x_3 , x_9 y x_4 las tres condiciones iniciales que permiten realizar el ajuste con el error más bajo.

Seleccionamos entonces los resultados para x_1 y x_3 como los parámetros óptimos encontrados para la rutina de *lsqcurvefit* y algoritmo genético, respectivamente. Al emplear la ecuación (2.4) para el cálculo del número básico de reproducción con dichos resultados se obtiene que para Cali dicha cantidad es $\mathcal{R}_0 = 1,59$ y $\mathcal{R}_0 = 1,43$ con los respectivos métodos empleados.

Capítulo 4

Conclusiones

En la práctica, como era de esperarse, aplicar el método de mínimos cuadrados no lineales aplicando el algoritmo genético requiere un mayor tiempo de ejecución que al aplicar la rutina de *lsqcurvefit*. Esto se debe a que al aplicar el algoritmo genético no se tiene en consideración la diferenciabilidad de la función a ajustar y hace uso solamente del conjunto de datos que se tienen, mientras que la rutina de *lsqcurvefit* utiliza el algoritmo de Levenberg-Marquardt que optimiza gran parte de los cálculos al hacer uso de la matriz Jacobiana de la función y una aproximación de su matriz Hessiana. Otra razón para la gran diferencia de tiempos de ejecución entre los algoritmos empleados es que gran parte de la elaboración del algoritmo genético se realiza sin paquetes ya optimizados de Matlab que cumplan con funciones específicas y permitan realizar la gran cantidad de cálculos ejecutados por el programa durante el bucle al cual se debe someter. Sin embargo, por ambos métodos empleados se obtienen cantidades muy similares para los parámetros estimados y, aunque la rutina de *lsqcurvefit* haya presentado un menor error cuadrático, ambos permitieron cumplir con el objetivo principal de este trabajo de grado, ayudando a su vez a ampliar los conocimientos sobre la programación y aplicación de estos métodos de estimación de parámetros.

Los resultados para los parámetros estimados permitirán analizar la dinámica de la enfermedad de tuberculosis en Cali con el modelo estudiado y así realizar predicciones sobre el comportamiento de esta en la ciudad, pues según los errores cuadráticos obtenidos por ambos métodos de estimación de parámetros aplicados y las gráficas de ajuste realizadas, los valores encontrados aproximan bastante bien el modelo al comportamiento real de la enfermedad. Puede decirse entonces de los valores encontrados que el tratamiento empleado es efectivo e influye de una manera significativa en el control y reducción de la infecciosidad de la enfermedad (ϕ), al igual que la tasa de reclutamiento para tratamiento (γ) que presenta un valor alto comparado con los demás parámetros estimados. Sin embargo, la alta tasa de desarrollo lento de la enfermedad (ϵ) y la baja tasa de transmisión de la tuberculosis (β) dan lugar a la interpretación de que la enfermedad puede persistir en la población por un periodo de tiempo prolongado.

Por otra parte, según los resultados obtenidos por ambos métodos, se deben establecer medidas de control para la enfermedad de TB en la ciudad de Cali, esto lo indica el número básico de reproducción $\mathcal{R}_0 > 1$, pues dice que una persona infectada que se encuentra en una población susceptible puede contagiar alrededor de tres personas cada dos meses, lo cual indica que el equilibrio endémico es el asintóticamente estable; es decir, que la enfermedad va a persistir en la región.

Bibliografía

- [1] <https://www.paho.org/es/temas/tuberculosis>.
- [2] <https://www.who.int/es/news-room/fact-sheets/detail/tuberculosis>.
- [3] <https://www.minsalud.gov.co/salud/publica/PET/Paginas/Tuberculosis.aspx>.
- [4] C. Agudelo, *Guías de promoción de la salud y prevención de enfermedades en la salud pública*. Bogotá, Colombia: Scripto, 2007.
- [5] E. Barrios-Rivera, H. E. Bastidas-Santacruz, C. A. Ramirez-Bernate y O. Vasilieva, “A synthesized model of tuberculosis transmission featuring treatment abandonment,” *Mathematical Biosciences and Engineering*, vol. 19, n.º 11, págs. 10 882-10 914, 2022.
- [6] N. Chitnis, J. M. Hyman y J. M. Cushing, “Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model,” *Bulletin of mathematical biology*, vol. 70, págs. 1272-1296, 2008.
- [7] S. Choi y E. Jung, “Optimal tuberculosis prevention and control strategy from a mathematical model based on real data,” *Bulletin of mathematical biology*, vol. 76, n.º 7, págs. 1566-1589, 2014.
- [8] L. Garcia Rovira et al., “Modelos matemáticos compartimentales en epidemiología,” 2017.
- [9] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Company, 1989.
- [10] Instituto Nacional de Salud, *Tuberculosis, Colombia, 2019*, https://www.ins.gov.co/buscador-eventos/Informesdeevento/TUBERCULOSIS_2019.pdf, 2019.
- [11] Instituto Nacional de Salud, *Comportamiento de la vigilancia de Tuberculosis, Colombia, 2020*, https://www.ins.gov.co/buscador-eventos/BoletinEpidemiologico/2021_Boletin_epidemiologico_semana_11.pdf, mar. de 2021.
- [12] D. R. Kincaid y E. W. Cheney, *Numerical analysis: mathematics of scientific computing*. American Mathematical Soc., 2009.
- [13] M. Martcheva, *An introduction to mathematical epidemiology*, 2015.
- [14] A. Rigalli, M. C. Aguirre, M. Armendáriz y G. Casiraga, *Formulación de modelos matemáticos de fenómenos biológicos*, 2003.
- [15] P. Van den Driessche y J. Watmough, “Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission,” *Mathematical biosciences*, vol. 180, n.º 1-2, págs. 29-48, 2002.
- [16] D. D. Wackerly, W. Mendenhall III y R. L. Scheaffer, *Estadística matemática con aplicaciones*. México: Cengage Learning, 2008.

- [17] G. P. Wormser y B. Pourbohloul, *Modeling Infectious Diseases in Humans and Animals By Matthew James Keeling and Pejman Rohani Princeton, NJ: Princeton University Press, 2008. 408 pp., Illustrated. 65,00(hardcover)*, 2008.
- [18] Y. Yang, J. Li, Z. Ma y L. Liu, “Global stability of two models with incomplete treatment for tuberculosis,” *Chaos, Solitons & Fractals*, vol. 43, n.º 1-12, págs. 79-85, 2010.
- [19] Y. Yang, S. Tang, X. Ren, H. Zhao y C. Guo, “Global stability and optimal control for a tuberculosis model with vaccination and treatment,” *Discrete & Continuous Dynamical Systems-B*, vol. 21, n.º 3, pág. 1009, 2016.