

**SIMULACIÓN Y ESTIMACIÓN DE PARÁMETROS DE MODELOS DE
ENFERMEDADES INFECCIOSAS**

GERMAN ANDRES OLIVEROS PATIÑO

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISCOQUÍMICAS
ESCUELA DE INGENIERÍA QUÍMICA**

2009

**SIMULACIÓN Y ESTIMACIÓN DE PARÁMETROS DE MODELOS DE
ENFERMEDADES INFECCIOSAS**

GERMAN ANDRES OLIVEROS PATIÑO

**Trabajo de grado para optar el título de
Ingeniero Químico**

**Director:
Dr. Carl D. Laird
Texas A&M University**

**Revisor de texto:
Profesor Crisóstomo Barajas
UIS**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISCOQUÍMICAS
ESCUELA DE INGENIERÍA QUÍMICA**

2009

AGRADECIMIENTOS

A la Universidad Industrial de Santander, mi alma mater

Al departamento de Ingeniería Química de Texas A&M y de la UIS por brindarme la oportunidad de realizar este proyecto de investigación.

Al Doctor Carl D. Laird, mi tutor y director de este trabajo, por su entrega a mi proyecto y por enseñarme a ser un buen investigador mediante su ejemplo.

A mi equipo de trabajo en Texas A&M: A James Young, por solucionar mis innumerables dudas, a George Abbott, Ethan Demeter, Derrick Thomas, Ahmed Rabie, Yu Zhu, Sean Legg y Chris Fain, el mejor equipo de investigación y los mejor colegas que pude haber tenido.

DEDICATORIA

A mis papás, las personas más incondicionales y entregadas que he conocido, por su constante preocupación por mi formación personal y profesional.

A Mónica, por su gran amor, su paciencia y la felicidad que me ha brindado

A mis amigos del colegio y de la universidad, por ser mi permanente fuente de alegría

RESUMEN

Título: SIMULACIÓN Y ESTIMACIÓN DE PARÁMETROS DE MODELOS DE ENFERMEDADES INFECCIOSAS*

Autor: German Andrés Oliveros Patiño**

Palabras clave: modelo de compartimientos, modelo SIR, sarampión, tasa de transmisión, epidemiología matemática, optimización

Una descripción detallada de modelos dinámicos que describen la evolución y proliferación de enfermedades infecciosas fue estudiada con el propósito de validar uno de éstos modelos, llamado SIR, utilizando datos reales. Se utilizaron los casos reportados de sarampión en Londres en años previos a la introducción de la vacunación masiva para obtener los parámetros importantes del sistema, como lo son la tasa efectiva de contactos y el número de susceptibles así como el ajuste del modelo a los datos reales. Para lograr este fin, se formuló un problema de optimización no lineal, donde el objetivo consistió en minimizar el error entre los datos predichos por el modelo y los datos reportados sujeto a las ecuaciones que describen el modelo.

Se encontró que la causa principal que explica el fenómeno cíclico del sarampión es el calendario académico de los colegios en Londres, donde las temporadas de vacaciones (navidad, pascua y verano) muestran una tasa de contactos inferior en comparación con aquella que se presenta en la temporada de estudio. Los resultados obtenidos fueron comparados con estudios previos donde la tasa de contactos era obtenida utilizando un método diferente llamado TSIR y se comprobó que los resultados obtenidos en éste trabajo proporcionan una alternativa mas flexible y mas rigurosa que la utilizada éstos trabajos.

* Trabajo de Investigación realizado en Texas A&M University

** Facultad de Ingenierías Físicoquímicas, Escuela de Ingeniería Química, Director: Dr. Carl D. Laird, Lector: Crisóstomo Barajas Ferreira

ABSTRACT

Title: SIMULATION AND PARAMETER ESTIMATION OF INFECTIOUS DISEASE MODELS*

Author: German Andres Oliveros Patiño**

Keywords: Compartment models, SIR model, measles, transmission rate, mathematic epidemiology, optimization

A detailed description of dynamic models that describe the evolution and spread of infectious diseases has been studied with the objective of validating one of them, the SIR model, by trying to reproduce real data. Biweekly reported cases of measles in London during the pre-vaccination era were used to obtain the main parameters of the system, such as the transmission rate and the amount of susceptibles in the population, as well as the model fit to the real data. A nonlinear optimization formulation was developed where the objective function was the minimization of the error between values predicted by the model and those from real data and the constraints were the set of the SIR differential equations.

Results showed that the main factor that explained the cyclic pattern of measles is the school calendar in London, where holiday seasons such as Christmas, Eastern and Summer had a lower contact rate with respect to school period. These results were compared with previous studies done by other authors where they utilized an alternate method, the TSIR to achieve such patterns. It was demonstrated that results obtained in this work provide a more flexible and rigorous alternative to previous approaches regarding this matter.

* Research Project, Texas A&M University

** Chemical Engineering Department, Advisor: Dr. Carl D. Laird, Reviewer: Professor Crisóstomo Barajas Ferreira

TABLA DE CONTENIDO

1. FUNDAMENTO TEÓRICO	2
1.1. Epidemiología matemática	2
1.2. Microparásitos y etapas de la enfermedad	2
1.3. La tasa de reproducción básica	4
2. METODOLOGÍA	5
2.1 Estudio y análisis de los modelos	5
2.1.1 <i>El modelo MSEIR</i>	5
2.1.2 <i>El modelo SIR</i>	7
2.1.3 <i>Solución al modelo SIR</i>	12
2.1.4 <i>Período de permanencia en los compartimientos</i>	15
2.2 Caso de estudio: Sarampión en Londres	16
2.2.1 Estimación de parámetros	19
3. RESULTADOS Y ANÁLISIS	21
3.1. Tasa de contactos	21
3.1.1 <i>La tasa de transmisión anual $\beta(t)$</i>	21
3.1.2 <i>Efecto de ρ sobre la tasa de contactos</i>	22
3.1.3 <i>Comparación con estudios previos</i>	23
3.2. Ajuste del modelo	24
3.3. Fracción de susceptibles en la población	25
CONCLUSIONES	27
RECOMENDACIONES	28
REFERENCIAS	28

LISTA DE FIGURAS

Figura 1. Períodos de incubación latente y de generación.....	3
Figura 2. El modelo MSEIR.....	5
Figura 3. El modelo SIR.....	7
Figura 4. Solución del modelo SIR utilizando valores representativos del sarampión	12
Figura 5. Número de susceptibles con respecto al tiempo (en días).....	13
Figura 6. Número de infectados con respecto al tiempo (en días).....	13
Figura 7. Número de recuperados con respecto al tiempo (en días).....	13
Figura 8. Descenso “clásico” en el número de individuos con respecto al tiempo.....	14
Figura 9. Formulación mejorada de los períodos de permanencia.....	15
Figura 10. Distribución realista del tiempo de permanencia.....	16
Figura 11. Casos reportados de sarampión en Londres entre 1944 y 1964....	17
Figura 12. Esquema general de la metodología a implementar para obtener las propiedades epidemiológicas del sistema $I(t)$, $S(t)$, $\beta(t)$	18
Figura 13. Tasa de contactos en función del tiempo.....	21
Figura 14. Figura del aumento del parámetro de regularización sobre la tasa de contactos.....	23
Figura 15. Patrón de contactos obtenidos por diferentes autores.....	24
Figura 16. Ajuste del modelo a los datos reportados	24
Figura 17. Población susceptible durante el período de estudio.....	26

LISTA DE TABLAS

Tabla 1. Períodos de incubación, latentes e infecciosos de diferentes enfermedades en días.....	3
--	----------

INTRODUCCIÓN

El modelamiento de enfermedades infecciosas estudia la dinámica y los factores que causan la proliferación de una enfermedad cuando ésta invade una población en particular. El presente trabajo estudió un grupo de sencillos modelos, representados por un sistema de ecuaciones diferenciales, que describen adecuadamente el proceso de infección y de transmisión de la enfermedad mediante la clasificación de la población en diferentes clases epidemiológicas. El objetivo principal de este trabajo es probar la validez de uno de éstos modelos, el modelo SIR, mediante el uso de datos de incidencia de sarampión en Londres en la época previa a la introducción de la vacunación masiva. El “acoplamiento” del modelo SIR a los casos reportados de sarampión permitieron estimar los parámetros relevantes del sistema, como la tasa efectiva de contactos y la cantidad de individuos susceptibles, así como también permitieron calcular el ajuste del modelo a los casos reales reportados. La metodología empleada consistió en formular un problema de optimización donde el la función objetivo era la minimización del error entre los datos predichos por el modelo y aquellos reportados mediante el uso de software de programación. Se encontró una fuerte correlación entre el calendario académico en los colegios de Londres y la dinámica del sarampión en dicha región, donde los períodos de vacaciones mostraron una tasa de contactos notablemente inferior a aquella del período académico de los niños. La utilización de un modelo continuo para obtener estos resultados demuestra un importante avance en el estudio de la dinámica de la transmisión de esta enfermedad ya que el enfoque utilizado en trabajos previos carecía de flexibilidad y estaban limitados a ciertas condiciones, como el tiempo entre cada reporte de casos y la duración de la enfermedad.

El estudio sistemático de la propagación de enfermedades infecciosas es una poderosa herramienta para reproducir, analizar, controlar y prevenir eventuales brotes de éstas enfermedades.

1. FUNDAMENTO TEÓRICO

1.1 Epidemiología matemática

El uso de modelos matemáticos para describir el comportamiento dinámico de una enfermedad infecciosa ha sido estudiado desde 1760, cuando Daniel Bernoulli utilizó un método matemático para modelar la viruela.

Como se mencionó anteriormente, estos modelos son herramientas muy útiles para analizar y controlar la propagación de una enfermedad y pueden ser usados para comparar, planear, implementar, evaluar y optimizar diversos programas de detección, prevención y terapia. Más aún, los modelos detallados y cuidadosamente examinados pueden estimar futuros eventos y brindar información sobre los periodos óptimos de vacunación. Algunos modelos recientes incluyen factores como la inmunidad pasiva, pérdida gradual de la vacuna e inmunidad adquirida por la enfermedad, etapas de la infección, mezcla entre grupos sexuales (en el caso de enfermedades de transmisión sexual), cuarentena y quimioterapia [1]. En general, estos son los aspectos que estudia la epidemiología matemática.

1.2 Microparásitos y etapas de la enfermedad

En el transcurso de este trabajo, se estudiará la transmisión de enfermedades infecciosas a través de los microparásitos. Éstos parásitos se reproducen a grandes velocidades dentro del huésped y generalmente una vez que el huésped se ha recuperado de la infección, éste será inmune a una reinfección por un determinado período de tiempo. Los virus, las bacterias, así como otros parásitos fungales y protozoarios caen en esta categoría [2].

El método más utilizado y a la vez conveniente de modelar la dinámica de una enfermedad infecciosa, es a través del llamado modelo de compartimientos. En este modelo, cada compartimiento es caracterizado por un estado, específicamente, una etapa de la enfermedad. La suposición más importante de este modelo consiste en asumir que el grado o la severidad de la infección

no es tenida en cuenta [1], en otras palabras, el individuo está infectado o no lo está.

Previo a la formulación del modelo, es necesario definir los períodos de incubación, latente e infeccioso. De acuerdo a la literatura concerniente a las infecciones causadas por microparásitos [2], el período de incubación se define como el tiempo en el que ocurre la infección y la aparición de los síntomas. El tiempo entre la infección y el estado infeccioso (donde el individuo puede transmitir la enfermedad) se denomina el período latente. Finalmente, la suma del período infeccioso y latente, se denomina el período de generación.

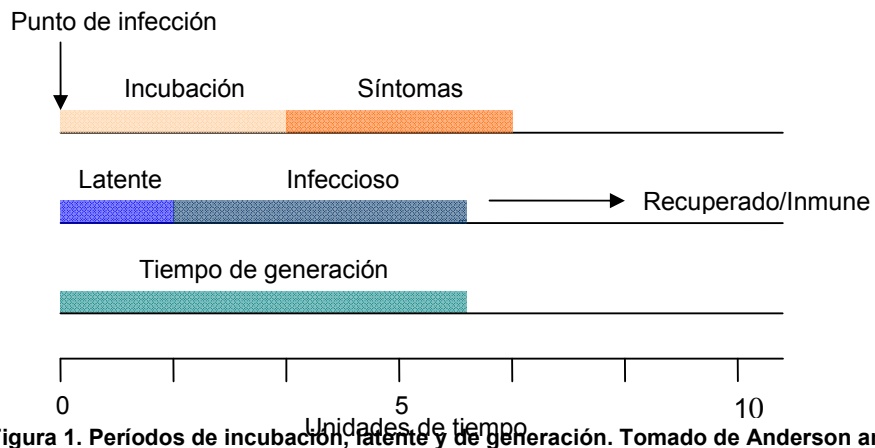


Figura 1. Períodos de incubación, latente y de generación. Tomado de Anderson and May [2].

ENFERMEDAD INFECCIOSA	PERÍODO DE INCUBACIÓN	PERÍODO LATENTE	PERÍODO INFECCIOSO
Sarampión	8 - 13	6 - 9	6 - 7
Rubeola	12 - 26	12 - 18	4 - 8
Escarlatina	2 - 3	1 - 2	14 - 21
Paperas	12 - 26	12 - 18	4 - 8
Varicela	13 - 17	8 - 12	10 - 11
Gripe	1 - 3	1 - 3	2 - 3
Viruela	10 - 15	8 - 11	2 - 3

Tabla 1. Períodos de incubación, latentes e infecciosos de diferentes enfermedades en días. Tomado de Anderson and May [2].

1.3 La tasa de reproducción básica

Uno de los parámetros más importantes usados en epidemiología es la tasa de reproducción básica R_0 , definida como el número promedio de nuevas infecciones causadas por un individuo infectado cuando es introducido en una población *completamente* susceptible. Para que una enfermedad persista R_0 deberá ser mayor a cero, de lo contrario la enfermedad no podrá invadir una población. Asumiendo una población totalmente homogénea (todos sus habitantes con las mismas características epidemiológicas y hábitos sociales), el número de infecciones es proporcional a la probabilidad de tener un contacto efectivo con un individuo susceptible. Es importante hacer la diferencia entre un contacto y un contacto efectivo. El último es aquel que permite transmitir la enfermedad a otra persona.

De acuerdo a esto, la tasa de reproducción *efectiva* se define como:

$$R = R_0 x$$

Donde x es la fracción de la población que es susceptible. Cuando $R=1$, entonces el sistema estará en equilibrio, lo que significa que cada infección producirá en promedio una infección secundaria. Este resultado nos permite llegar a una importante relación entre la tasa de reproducción básica y la fracción susceptible de una población en el equilibrio.

$$R_0 x = 1$$

De nuevo, esta relación es válida en poblaciones no heterogéneas.

2. METODOLOGÍA

2.1. Estudio y análisis de los modelos

2.1.1 El modelo MSEIR

El modelo MSEIR clasifica a los individuos de una población en cinco categorías: Pasivamente Inmunes (**M**), Susceptibles (**S**), Expuestos (**E**), Infectados (**I**) y Recuperados (**R**).

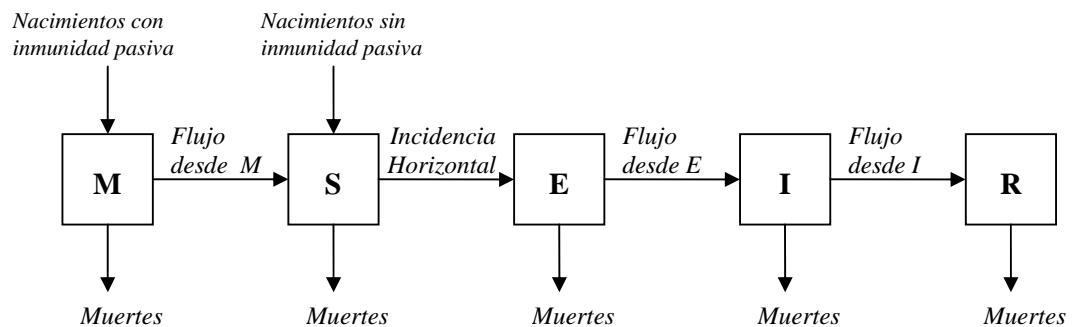


Figura 2. El modelo MSEIR

En el esquema de arriba es posible sintetizar el comportamiento y evolución de una enfermedad infecciosa. Por ejemplo, cuando una madre (**M**) que ha sido infectada con una enfermedad que confiere inmunidad permanente da a luz a un individuo, algunos de sus anticuerpos serán transferidos hacia el infante y aquel individuo será pasivamente inmune al momento del nacimiento. Por el contrario si una madre nunca ha estado expuesta a la enfermedad, el recién nacido será vulnerable a la enfermedad y será clasificado como susceptible (**S**). Cuando el individuo susceptible (**S**) entra en contacto efectivo con un individuo infectado (**I**), entonces éste será “transferido” al compartimento de población expuesta (**E**) y tendrá la infección pero no la habilidad para transmitirla a otras personas, en otras palabras, están infectados más no son infecciosos (período latente). El movimiento que describe este flujo se denomina incidencia horizontal y será descrito más adelante en el trabajo.

Después de que el período latente ha finalizado, el individuo infectado entra en el compartimiento de la clase infectada (**I**) y tiene ahora la capacidad para infectar a otras personas dentro de la población. El individuo permanecerá en este compartimiento durante el tiempo que dure la infección (ver Tabla 1). Después de que éste período ha finalizado, el individuo infectado es transferido al compartimiento de la clase recuperada (**R**), y, dependiendo si la enfermedad confiere inmunidad permanente o no, el individuo permanecerá en esta categoría para siempre o regresará de nuevo al compartimiento de la clase susceptible (**S**).

En este esquema existe un flujo de salida en cada compartimiento que representa las muertes de individuos en cada compartimiento. Usualmente esta clase de flujos se usan para modelar brotes endémicos, es decir, aquellos que afectan de manera continua una población o zona geográfica en particular; contrario a los brotes epidémicos, donde la enfermedad no persiste por mucho tiempo (generalmente menos de un año). Factores como los nacimientos y las muertes en una población son característicos de los eventos endémicos.

Como es posible deducir, hay miles de combinaciones [3] que pueden ser construidas a partir de esta clase de modelos. El MSEIR, es sólo uno de los varios modelos que podemos obtener: el SIR (sin período latente ni inmunidad pasiva), SIS (no hay inmunidad después de la infección), MSEIRS (sin inmunidad permanente).

Es importante tener en cuenta que cada enfermedad tiene propiedades epidemiológicas diferentes, como por ejemplo, el tiempo en el que cada individuo permanece en cada compartimiento. Otro aspecto fundamental es la relación que describe el movimiento de un compartimiento a otro. Estas relaciones son implementadas en un conjunto de ecuaciones diferenciales, como las que se mostrarán a continuación.

2.1.2 El modelo SIR

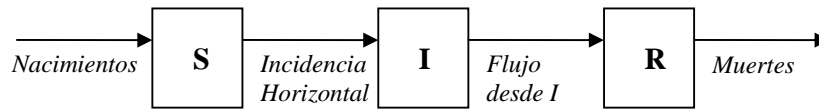


Figura 3. El modelo SIR

El uso de modelo SIR es ampliamente utilizado para describir el comportamiento dinámico de enfermedades infantiles ya que posterior a la infección, el niño o niña adquiere inmunidad vitalicia. Dentro de las enfermedades que se ajustan a este modelo, se encuentra el sarampión [4], que será el caso de estudio posterior al desarrollo del modelo.

Las ecuaciones que constituyen el modelo SIR son ecuaciones de conservación aplicadas a cada compartimiento. Análogo a la masa, el término de generación/desaparición no se aplica en una población de individuos, así que las expresiones resultantes para cada compartimiento serían:

Susceptibles “acumulados” = Infectados que entran – Infectados que salen

Infectados “acumulados” = Infectados que entran – Infectados que salen

Recuperados “acumulados” = Recuperados que entran – Recuperados que salen

El siguiente paso es expresar en una manera detallada cada uno de estos términos. Para un modelo SIR endémico, los nacimientos son el único flujo de entrada al compartimiento de susceptibles (**S**). Si se asume que la tasa de natalidad y la tasa de mortalidad son iguales (como sucede en la mayoría de países desarrollados) para que la población N permanezca constante, entonces se puede utilizar un mismo parámetro para esta tasa, denominada en este caso μ . Por ejemplo, si la edad promedio de vida en una población es de 75 años, entonces $1/\mu$ será 75 años y la tasa de natalidad y mortalidad promedio será de $1/75 \text{ años}^{-1}$. Por consiguiente, el número de nacimientos por unidad de tiempo en una población constante es μN .

Teniendo descritos los términos de la dinámica vital del sistema, se procede a describir el término de *incidencia horizontal*. Este término expresa la tasa a la

cual los susceptibles adquieren la infección y se desplazan al comportamiento de los infectados (I). Sea Beta (β) el número de contactos efectivos que una persona tiene por unidad de tiempo e $I(t)$ el número de infectados en el tiempo t , entonces el número de contactos que una persona tiene con un individuo infectado es:

$$\beta \cdot \frac{I(t)}{N}$$

Si $S(t)$ es el número de susceptibles en una población en el tiempo t entonces el número de nuevos casos de la infección por unidad de tiempo debido a esos susceptibles será:

$$\beta \cdot \frac{I(t)}{N} \cdot S(t)$$

Esta expresión es que la que comúnmente se denomina *incidencia horizontal*. En epidemiología es importante distinguir entre los términos incidencia y prevalencia. El primero es el *número de nuevos casos en un determinado intervalo de tiempo* definido por la ecuación previamente deducida, mientras que prevalencia es el *número de casos en un punto en el tiempo*, esto es $I(t)$.

En esta derivación se asumió que el número de contactos no varía con el tiempo, sin embargo, en la mayoría de los casos este parámetro cambia de acuerdo a la estación o a la temporada del año. Por ejemplo, en las enfermedades infantiles, la tasa de contactos en el período de vacaciones es mucho menor que la tasa de contactos en el período académico.

La ecuación diferencial resultante para el compartimiento de los individuos susceptibles es:

$$\frac{dS(t)}{dt} = \mu \cdot N - \beta \cdot \frac{I(t)}{N} \cdot S(t)$$

$$S(0) \geq 0$$

Para realizar un análisis análogo al anterior para el compartimiento de los infectados (**I**) es necesario tener conocimiento del tiempo de duración de la infección. Sea $1/\gamma$ el período promedio de infección, entonces $\gamma I(t)$ es la tasa a la cual las personas infectadas salen de su compartimiento. La ecuación diferencial resultante sería:

$$\frac{dI(t)}{dt} = \beta \cdot \frac{I(t)}{N} \cdot S(t) - \gamma \cdot I(t)$$

$$I(0) \geq 0$$

Gamma (γ) es un parámetro equivalente a Mu (μ) y representa la tasa promedio a la cual el individuo se recupera. Es bastante común utilizar esta combinación de términos (multiplicación de una variable por un parámetro [1]), ya que resulta bastante conveniente para la solución del sistema, pero como se explicará más adelante, esta no es una descripción realista del proceso.

Debido a que en el modelo SIR se asume que el individuo una vez infectado adquiere inmunidad permanente después de haber sido infectado, los individuos que salen del compartimiento **R** no volverán de nuevo a ser susceptibles y solo saldrán del este compartimiento hasta su muerte, la ecuación resultante es:

$$\frac{dR(t)}{dt} = \gamma \cdot I(t) - \mu \cdot N(t)$$

$$R(0) \geq 0$$

O simplemente,

$$R(t) = N - S(t) - I(t)$$

Ya que la población permanece constante.

La mayoría de los parámetros epidemiológicos y demográficos como la tasa de mortalidad y de natalidad, tasas de recuperación y pérdida de inmunidad, pueden ser obtenidos directamente de datos y estudios apropiados. Por el otro

lado, la tasa de transmisión (incidencia horizontal) envuelve una gran cantidad de factores ambientales, sociales, biológicos que no se pueden obtener directamente.

Como todos los modelos que tratan de reproducir o representar fenómenos naturales, estos modelos de compartimientos tienen desventajas obvias y significativas. Todos estos modelos asumen una población homogénea (sin distribución de edades y con los mismos hábitos socioeconómicos). También asumen interacción homogénea entre los individuos de una población, lo que es bastante irreal. Por ejemplo, es sabido que los niños tienen una tasa de contacto mayor entre ellos que con los adultos, y que éstos últimos a su vez interactúan menos entre sí.

La gran ventaja que ofrecen estos modelos, es que nos permiten obtener y estudiar algunas propiedades epidemiológicas importantes de algún brote conservando cierto nivel de simplicidad.

2.1.3 Solución al modelo SIR

A continuación se presenta la solución de un modelo SIR con una cantidad inicial de susceptibles $S(0)$ igual a 99, una cantidad inicial de infectados $I(0)$ igual a 1 y ningún individuo recuperado. De esta manera la población total es de 100, asumiendo que la tasa de natalidad es igual a la tasa de mortalidad.

También se asumió que todos los recién nacidos son susceptibles a la enfermedad. Con respecto a las propiedades epidemiológicas se usaron valores característicos del sarampión. Se usó como tiempo de infección $1/\gamma$ la suma de los tiempos latentes e infecciosos (12 días, ver Tabla 1). El valor para la tasa de contacto se tomó como 2 contactos por día de acuerdo a la literatura [6] y el promedio de edad de vida se tomó como 75 años, que es la esperanza de vida en los países desarrollados. Es posible observar que inicialmente hay una disminución drástica en la cantidad de susceptibles y un incremento en la fuerza de la infección $(\beta/N)*I(t)$. Esto puede ser interpretado claramente como un sistema depredador-presa: La población infectada –el depredador- encuentra una superabundancia de susceptibles –presa- y la infección se

empieza a propagar hasta que los individuos infectados sobreexplotan la población susceptible, causando una disminución en la población infectada y un incremento continuo de la población recuperada.

El sarampión puede ser muy bien representado por el modelo SIR [4,13], así como también puede ser representado por el modelo SEIR (donde se hace la distinción entre el período latente y el infeccioso). La inmunidad permanente que confiere el sarampión es otra característica importante, y debido a que esta enfermedad no es mortal en los países desarrollados, todas las muertes serán muertes naturales, lo que es equivalente a incluir el flujo de muertes sólo en el compartimiento de los recuperados. El modelo también asume que todos los recién nacidos son susceptibles a contraer la enfermedad.

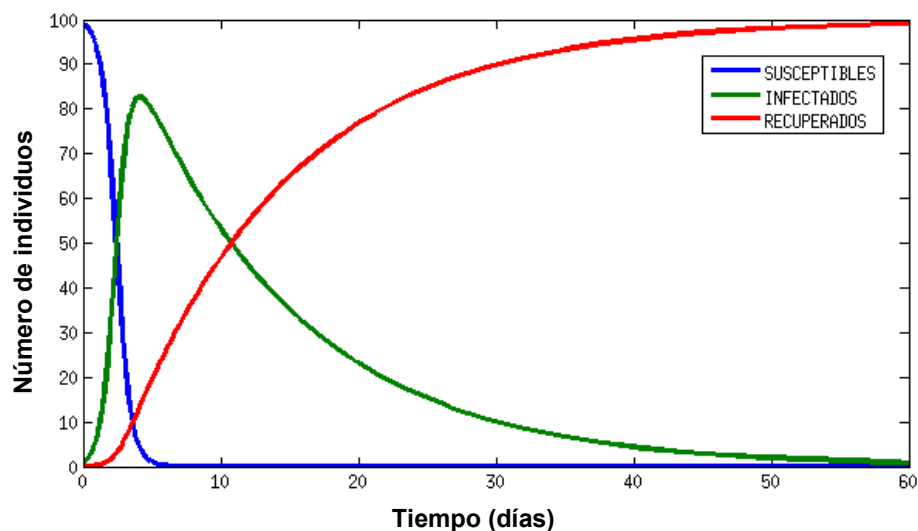


Figura 4. Solución del modelo SIR utilizando valores representativos del sarampión

Como ejemplo alternativo, ahora se muestra el comportamiento de lo que puede ocurrir en un escenario real. En la simulación anterior se utilizaron valores relativamente altos para el número de infectados, lo cual no es un escenario realista, ya que si esto sucediera, ninguna enfermedad persistiría en el tiempo.

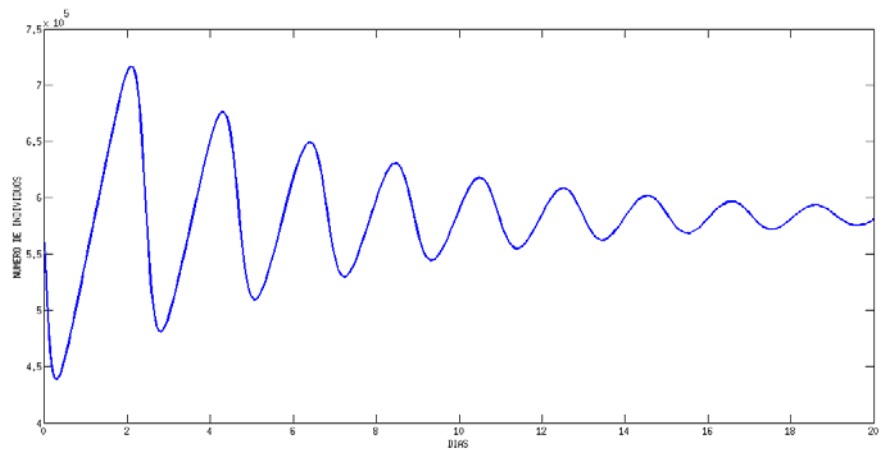


Figura 5. Número de susceptibles con respecto al tiempo (en días)

En el caso que se presenta a continuación, se escogieron valores similares a los que pueden presentarse en un brote real, en la cual, de una población total de aproximadamente 14 millones, el 4% de sus pobladores son susceptibles a contraer la enfermedad (aproximadamente unos 600 000), el 0,3% se encuentra infectado (aproximadamente 42 000), y el resto tienen inmunidad permanente (permanecen clasificados como Recuperados hasta su muerte). En este caso se puede observar que las variables epidemiológicas persisten en el tiempo. El parámetro que determina la persistencia y la desaparición de una enfermedad, es el número básico de reproducción.

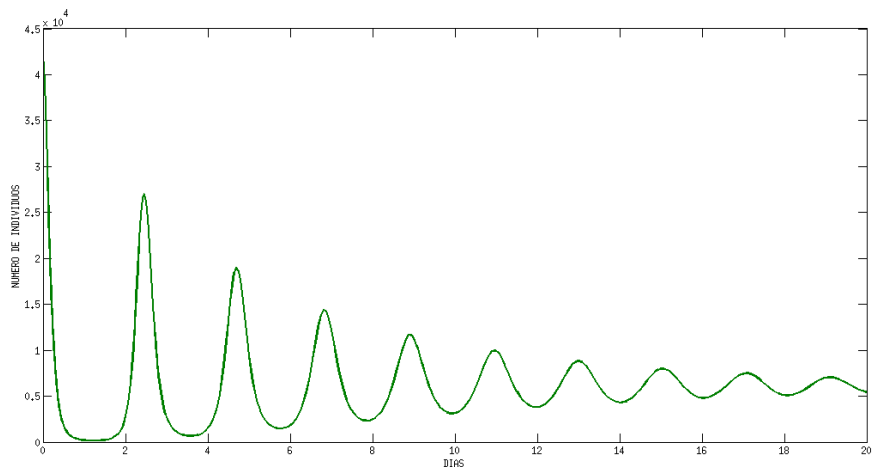


Figura 6. Número de infectados con respecto al tiempo (en días)

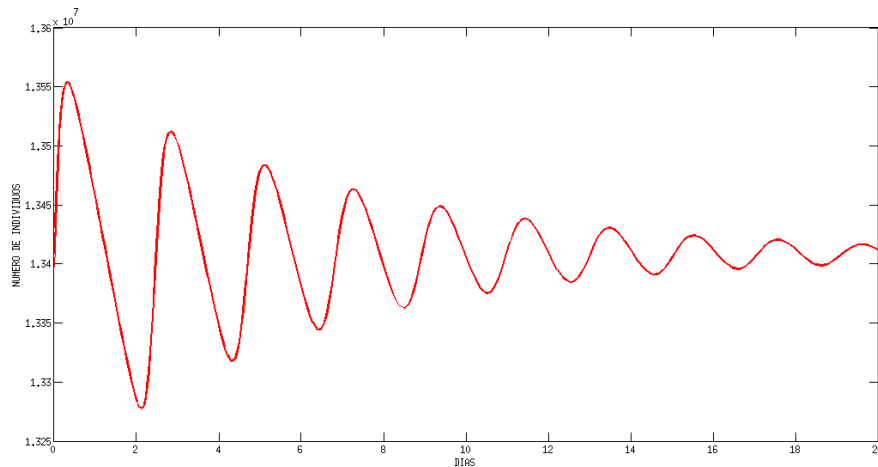


Figura 7. Número de recuperados con respecto al tiempo (en días)

2.1.4. Períodos de permanencia en los compartimientos

En el esquema clásico, el término que representa el tiempo de permanencia en un compartimiento tiene la forma $\alpha V(t)$ (por ejemplo, $\gamma I(t)$), donde α es una constante y $V(t)$ es cualquiera de las variables del sistema. Es posible analizar gráficamente la manera en la que los individuos salen de un compartimiento mediante la formulación y simulación de la siguiente ecuación diferencial.

$$\frac{dV(t)}{dt} = -\alpha * V(t)$$

$$V(0) \geq 0$$

Esta suposición del tiempo de permanencia implica que *la probabilidad de que un individuo salga del compartimiento es constante sin importar el tiempo que ha transcurrido desde que entro al compartimiento*. En el ejemplo de arriba se usó un valor inicial de 100 para la variable $V(t)$ y un período de permanencia de 1 semana (el valor de α sería 1 semanas⁻¹).

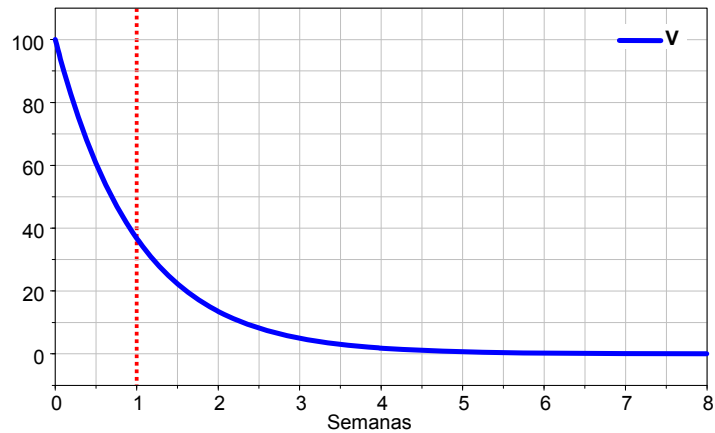


Figura 8. Descenso “clásico” en la cantidad de individuos con respecto al tiempo. El tiempo de permanencia $1/\alpha$ es 1 semana.

Esta es una suposición conveniente desde el punto de vista matemático, pero no es una descripción exacta del proceso. El comportamiento de esta disminución en el número de individuos debe ser tal, que al principio la cantidad de individuos saliendo del compartimiento es muy baja hasta el momento que alcanzan el tiempo promedio de permanencia, donde la tasa de individuos saliendo se incrementa drásticamente [9]. Después de este incremento la tasa vuelve a disminuir, generando así una distribución menos dispersa y más adecuada.

Una herramienta matemática es usada para generar esta clase de distribución y consiste en dividir el compartimiento en n “subcompartimientos” iguales como es mostrado en la figura de abajo. A medida que n se incrementa la distribución del período de permanencia se centra cada vez más alrededor del tiempo promedio (en este caso 1 semana) y si n se hace lo suficientemente grande todos los individuos permanecerán durante igual cantidad de tiempo en el compartimiento⁷.

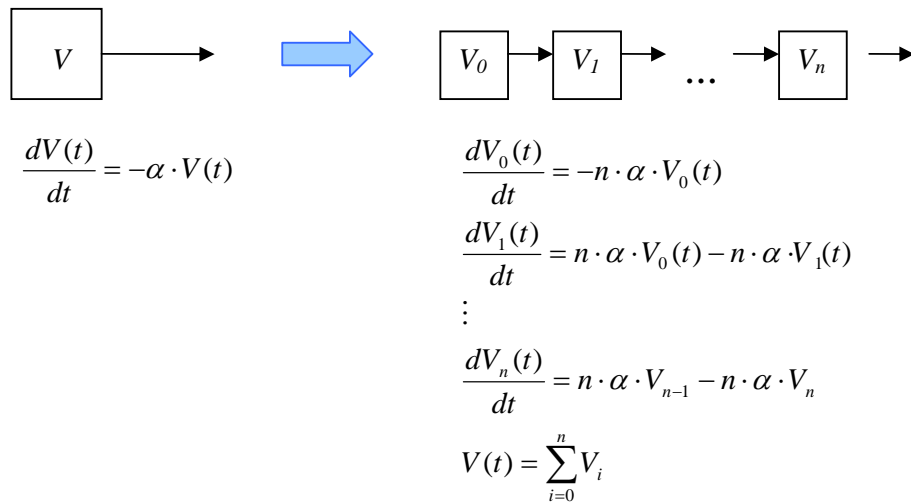


Figura 9. Formulación mejorada de los periodos de permanencia.

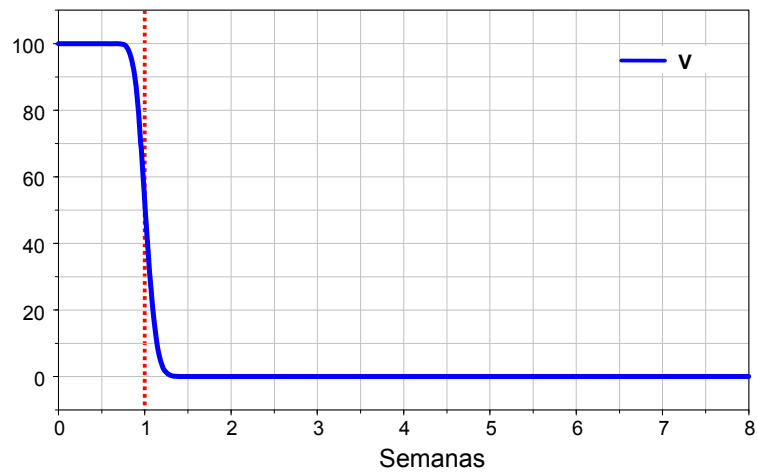


Figura 10. Distribución realista del tiempo de permanencia de los individuos en el compartimiento

2.2 Caso de estudio: Sarampión en Londres

Luego de presentados los conceptos básicos sobre la dinámica y el modelamiento de una enfermedad infecciosa, se realizará el estudio de un caso real. La información disponible tomada del *Registrars General Office* en

Inglaterra (disponible al público general) de los casos de sarampión reportados quincenalmente en Londres, Inglaterra (ver Figura 11) serán utilizados para obtener los parámetros relevantes del sistema, como lo son la cantidad inicial de susceptibles, el patrón anual de contactos así como el ajuste del modelo a los datos. Estos datos pertenecen a la época previa a la introducción masiva de los programas de vacunación (el efecto de la vacunación tiene una influencia importante en el modelamiento de enfermedades y no será estudiado en este trabajo) y cubren los años entre 1944 a 1966. De acuerdo a otros estudios realizados paralelamente a éste [8], aproximadamente el 47% de los casos de sarampión fueron reportados a las entidades de salud. Los datos de tasa de nacimientos, población y duración de la enfermedad también se conocen ya que están ampliamente documentados en el *Registrars General Office*. Como se puede apreciar en la figura de abajo, éste fenómeno no es puramente aleatorio. Existe un ciclo que persiste a través del tiempo y patrones que explican el comportamiento de esta enfermedad, debido a que el sarampión ocurre en la población infantil (en niños de aproximadamente 5 años [2]), los períodos escolares es probablemente una de las causas de ésta dinámica. Varios estudios han empleado técnicas diferentes y han obtenido resultados similares en lo que respecta a la influencia de la temporada escolar [10,11]. Lo que se pretende en este trabajo es utilizar una metodología más flexible para el estudio de ésta clase de problemas basados en datos.

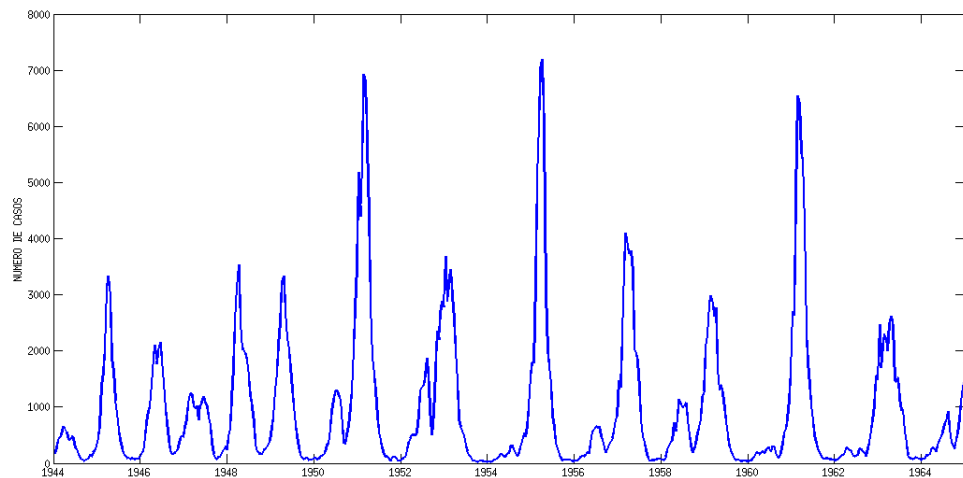


Figura 11. Casos reportados de sarampión reportados en Londres entre 1944 y 1964

El objetivo principal de este trabajo es probar la validez del modelo SIR cuando se aplica a un caso real. Acoplado éste a un problema de optimización, donde se busca minimizar el error entre los datos registrados y aquellos predichos por el modelo, utilizando la información demográfica y epidemiológica del sistema se logra este propósito. Las variables principales de este problema son $S(t)$, $I(t)$ y $\beta(t)$.

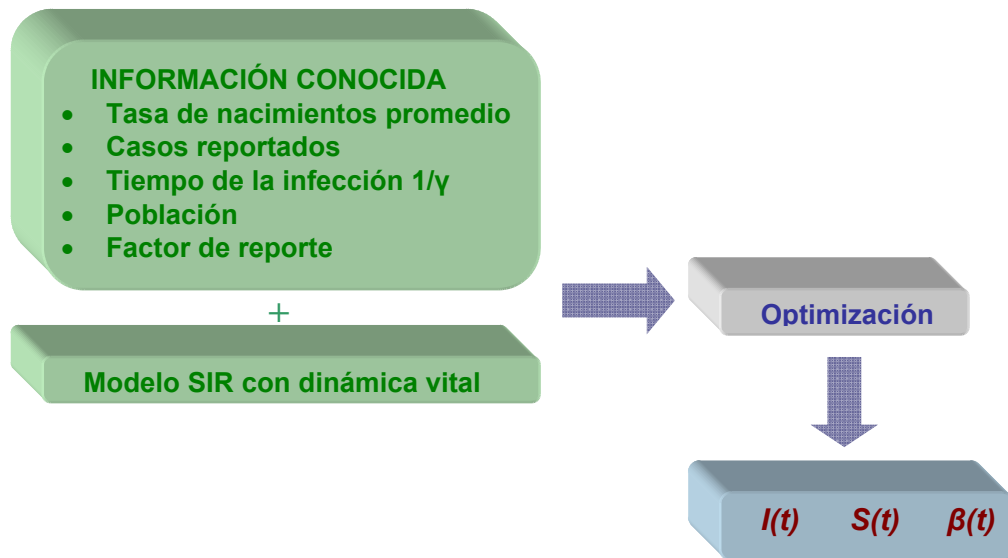


Figura 12. Esquema general de la metodología a implementar para obtener las propiedades epidemiológicas del sistema $I(t)$, $S(t)$, $\beta(t)$

Previo a la formulación del problema de optimización es necesario hacer las suposiciones principales bajo las cuales se va a trabajar. Como se mencionó anteriormente en este trabajo, el modelo SIR asume que la población interactúa de manera homogénea. Al estudiar el sarampión, se puede aprovechar esta propiedad, ya que la mayoría de los casos reportados ocurren en un mismo grupo de edad (niños entre 4 y 6 años).

La siguiente suposición está relacionada con los aspectos demográficos de Londres para esa época. Generalmente en los países desarrollados la población no varían mucho con respecto al tiempo, así se asumieron valores promedio entre 1944 y 1964 que para las tasa de natalidad y mortalidad valores y para la población se asumió que era aquella de 1944 y que permanecía

constante (3 249 440). Estas suposiciones facilitan considerablemente la resolución del problema en el software utilizado.

2.2.1 Estimación de parámetros

La formulación del problema de optimización es la siguiente

$$\min \int_0^T \left(R \cdot \text{phi}^*(t) - \text{phi}_r(t) \right)^2 dt$$

sujeto a :

$$\frac{dS(t)}{dt} = -\frac{\beta(t)}{N} \cdot I(t) \cdot S(t) + \mu N$$

$$\frac{dI(t)}{dt} = \frac{\beta(t)}{N} \cdot I(t) \cdot S(t) - \gamma \cdot I(t)$$

$$\text{phi}^*(t) = \int_{t_n}^{t_{n+1}} \frac{\beta(t)}{N} \cdot I(t) \cdot S(t) dt$$

$$-\rho \leq \frac{d^2 \beta(t)}{dt^2} \leq \rho$$

La función objetivo del problema de optimización es el error entre el número de casos reportados $\text{phi}_r(t)$ y aquellos predichos por el modelo SIR $\text{phi}^*(t)$ entre 1944 (tiempo 0) y 1964 (tiempo T). El factor de reporte es presentado como R, y es igual a 0,47 (47% de los casos ocurridos fueron reportados).

Esta función objetivo tiene 4 restricciones. Las primeras dos restricciones son las ecuaciones diferenciales del modelo SIR (la ecuación para R no se necesita ya que se asumió que la población se mantenía constante así que R es

simplemente $N-S(t)-I(t)$). En la primera ecuación μN es el número de nacimientos promedio durante el período de estudio (aproximadamente 2000 nacimientos quincenales).

La tercera restricción es el número de nuevos casos predichos por el modelo. Es importante tener en cuenta que el número de casos reportados está distribuido discretamente (cada 15 días) mientras que nuestro modelo SIR es un modelo continuo, así que es necesario calcular el número de nuevos casos predichos por el modelo mediante la integración de su función en intervalos quincenales, representado por los límites de la integral t_n y t_{n+1} .

El patrón de contactos anual es una propiedad característica de una población y se mantiene constante durante los 21 años de estudio. Esto es equivalente a:

$$\beta(t) = \beta(t \bmod 26)$$

Ya que las unidades de tiempo de la formulación serán en quincenas (hay 26 quincenas en 1 año). Esto también ayudará a facilitar el análisis y la comparación con otros estudios.

Finalmente, para evitar que $\beta(t)$ se comporte de una manera no suavizada, es necesario regularizar la función; esto se puede lograr utilizando uno tipo de regularización muy sencillo, y es restringir la segunda derivada de $\beta(t)$ mediante el uso de un límite positivo ρ y uno negativo $-\rho$. De esta forma se restringe la rapidez de cambio de la pendiente de la tasa de contactos; entre mas pequeños sean estos números (mas cercanos a cero), mas suave será la función. El principal inconveniente de esta formulación es el término no lineal representado por el término de incidencia $(\beta(t)/N)S(t)I(t)$, que es el término que brinda al sistema la mayor variabilidad y complejidad. Cuando existen términos no lineales en un problema de optimización, es crítico inicializar las variables de una manera apropiada para que el programa de computador lo pueda solucionar.

3. RESULTADOS Y ANÁLISIS

3.1. Tasa de contactos

3.1.1. La tasa de transmisión anual $\beta(t)$

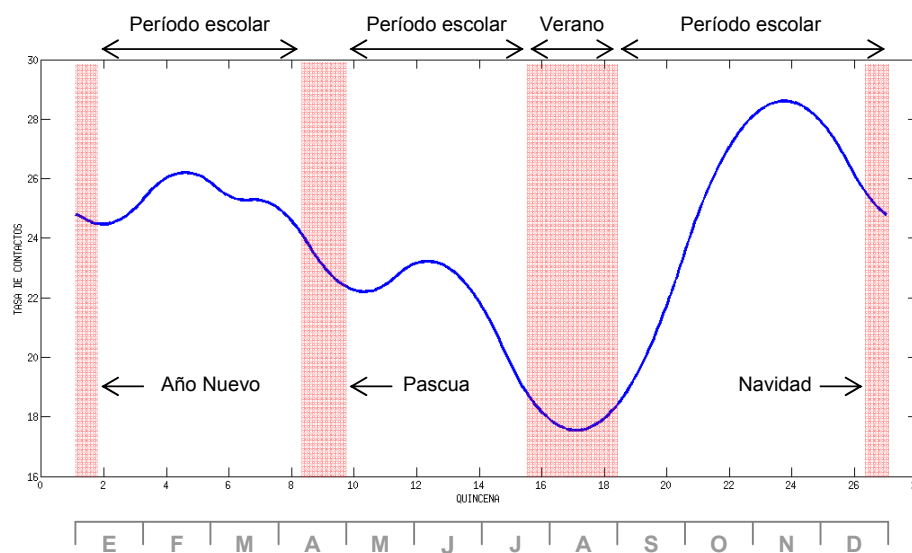


Figura 13. Tasa de contactos en función del tiempo (en quincenas y en meses)

La tasa de contactos es el factor principal que explica la dinámica del sarampión. Los valores de la tasa de contactos se encuentran entre aproximadamente 8 contactos por quincena, hasta aproximadamente 29 contactos por quincena, con un valor promedio de 23,5. En esta gráfica se pueden apreciar tres regiones importantes donde existe una disminución en la tasa de contactos. Estas disminuciones corresponden a los períodos de vacaciones de acuerdo al calendario escolar de Londres (y de toda Inglaterra): Navidad (quincena 1 y 26), Pascua (quincena 8, mitad de abril) y las vacaciones de verano (quincenas 15 hasta 18, mitad de julio hasta principios de septiembre). Estos son los períodos del año donde los niños pasan menos tiempo con sus contemporáneos y por lo tanto la transmisión de la enfermedad disminuye. Por el otro lado, el incremento mas pronunciado en la tasa de transmisión ocurre cuando el verano acaba y los niños vuelven al colegio causando un nuevo “flujo” de niños susceptibles al sistema (el calendario

escolar en Londres finaliza con el comienzo del verano). El comportamiento fluctuante de la tasa de contactos durante el período escolar es un reflejo de la suposición que se hizo sobre la población, donde se asumió que ésta era homogénea. Aún así, como la figura lo confirma es una muy buena aproximación. Esto también indica que las escuelas son la principal, pero no la única, fuente de adquisición de la enfermedad (de lo contrario la tasa de transmisión durante el calendario escolar sería constante).

3.1.2. Efecto de ρ sobre la tasa de contactos

A continuación se presenta la tasa de contactos utilizando un valor diferente al presentado anteriormente. De esta forma se analiza que a pesar de que la suavidad de la función varía, el patrón estacional se conserva, es decir, se siguen reproduciendo adecuadamente los períodos escolares. El valor utilizado en este caso el valor para ρ es de 7, en comparación con un valor de 1 utilizado en la Figura 13. Esto le permite a la pendiente de la función $\beta(t)$ variar con mayor libertad. Valores de ρ de hasta 13 pueden aún obtener este mismo patrón pero debido a las oscilaciones no permiten distinguir tan claramente los períodos académicos y de vacaciones.

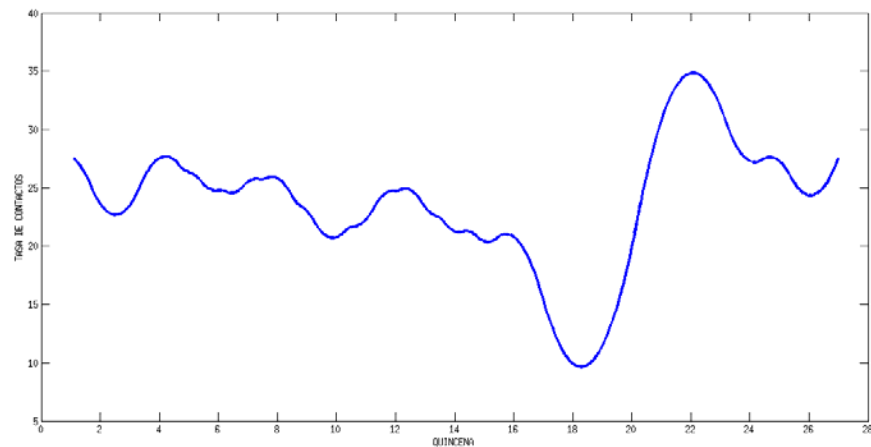


Figura 14. Efecto del aumento del parámetro de regularización sobre la tasa de contactos.

3.1.3. Comparación con estudios previos

Con la finalidad de probar la validez del estudio realizado sobre el modelamiento del sarampión en Londres, se muestra la tasa de contactos obtenidas por autores diferentes de manera sucinta.

En la figura de la derecha se muestra el resultado obtenido por B.F Finkenstadt y B.T Grenfell [4]. La metodología utilizada en sus estudios consistía en aprovechar la similitud entre el tiempo que dura la infección (14 días) y el tiempo entre cada reporte de la enfermedad (cada quincena), es por esta razón que la tasa de contactos no tiene un comportamiento suavizado. El modelo utilizado por estos autores es comúnmente llamado TSIR (*time-series SIR*) [4]. La figura 17, muestra mas detalladamente el período de las vacaciones en comparación con las otras investigaciones, donde ocurren descensos en la tasa de contactos después de Enero, donde en teoría los niños ya deberían estar en temporada escolar, mientras que en la obtenida en este estudio muestra una disminución en la tasa de transmisión terminando el año (Navidad) continuando en los primeros días del año (Año Nuevo). Otro enfoque desarrollado por Cauchemez y Ferguson[11] realizan un análisis estadístico junto con el modelo TSIR para obtener este patrón de contactos. De nuevo, sus resultados son consistentes con los obtenidos en este estudio.

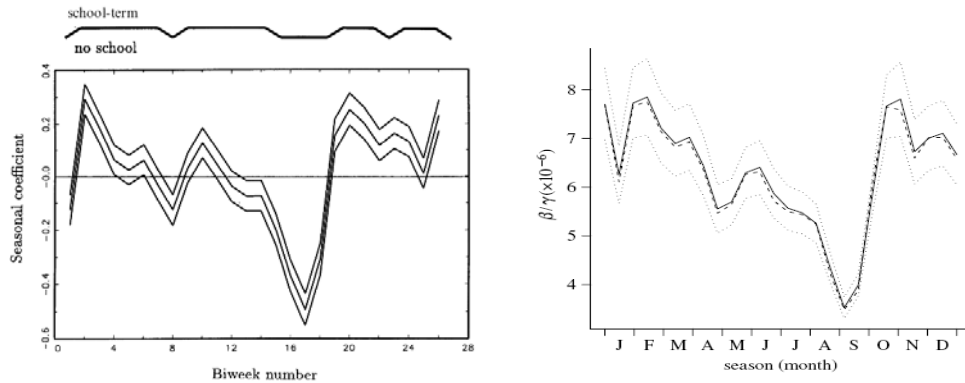


Figura 15. Patrón de contactos obtenidos por diferentes autores.

3.2. Ajuste del modelo

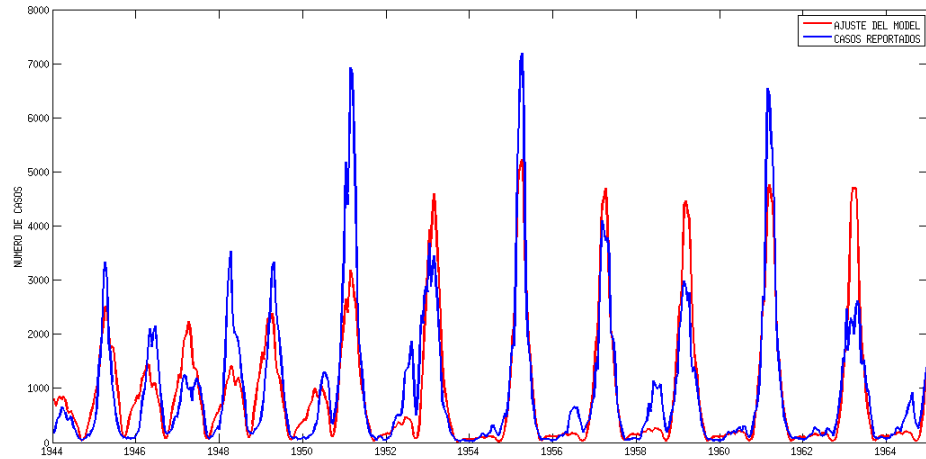


Figura 16. Ajuste del modelo a los datos reportados

El aspecto mas importante al analizar el ajuste del modelo a los datos es el hecho de que un modelo completamente determinístico nunca se ajustará perfectamente a un fenómeno donde los aspectos estocásticos inherentes a la población juegan un papel fundamental en el patrón cíclico del sarampión así como las suposiciones hechas sobre el modelo utilizado. Sin embargo, como se puede observar en la figura, el modelo es capaz de reproducir y acoplarse a los brotes de una manera precisa durante todo el período de estudio. La inclusión de factores estocásticos, así como la clasificación de las clases epidemiológicas de acuerdo a la edad (modelo SIR descrito por ecuaciones diferenciales parciales con variables que cambian con el tiempo y la edad [12]) pueden ayudar a reproducir un mejor patrón de contactos, así como la inclusión de una mejor distribución del tiempo de la enfermedad (debido al aumento de la dificultad en la programación que esto generaría, no se implementó la distribución realista en la formulación del problema de optimización). En un futuro el objetivo no sólo será reproducir los datos con mayor exactitud sino poder *predecir* un eventual brote en una población.

3.3. Fracción de susceptibles en la población

En la figura de debajo se muestra la cantidad de susceptibles durante los años de estudio. Se encontró que en 1944 el 4,6% de la población era susceptible al sarampión. Si se tiene en cuenta la ecuación que relaciona el número de infecciones secundarias que un individuo infectado transmite y la fracción de susceptibles en una población en el equilibrio

$$1 = R_0 x$$

Y en adición se tiene en cuenta que $\beta(t)$ es el número de contactos efectivos (suficientes para adquirir la infección) que una persona tiene por unidad de tiempo, entonces, en teóricamente, la fracción de individuos susceptibles en el caso del sarampión en Londres debe ser

$$\frac{1}{\bar{\beta}} = x$$

Donde $\bar{\beta}$ es el valor promedio de la tasa de contactos. Utilizando esta aproximación se encuentra que

$$\frac{1}{23,5} = x = 0,0425 = 4,25\%$$

Valor bastante cercano al obtenido en la simulación.

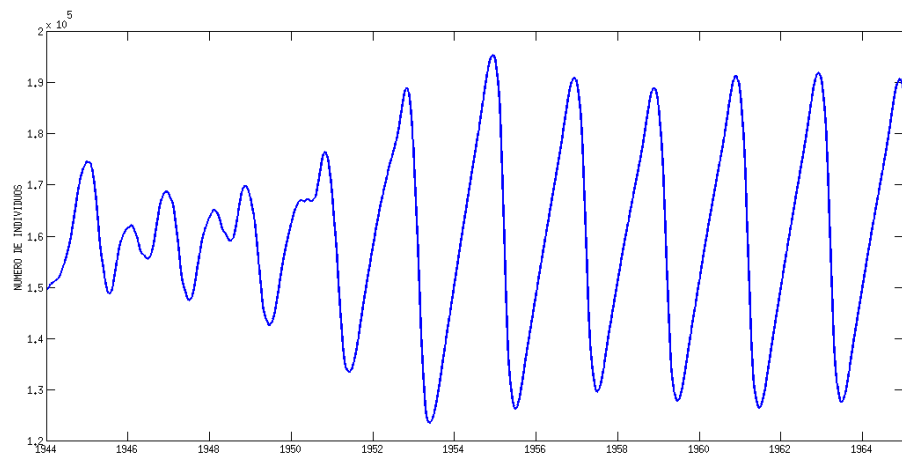


Figura 17. Cantidad de susceptibles durante el período de estudio.

CONCLUSIONES

En este trabajo se presentaron los fundamentos del modelamiento de enfermedades infecciosas utilizando un modelo de compartimientos y se mostraron las posibles mejoras para capturar de una manera más acorde y realista la evolución de una infección dentro de una población.

La validación del modelo continuo SIR con dinámica vital mediante su acoplación a datos reales representa un avance significativo en el estudio del modelamiento de enfermedades infecciosas en poblaciones infantiles. El modelo continuo brinda flexibilidad con respecto a los métodos utilizados en las investigaciones relacionadas con el sarampión en Londres en la época previa a la introducción masiva, ya que no depende de la discretización del período de reporte ni de la similitud que este debe tener con el período de la infección, que era la forma en que el problema estaba siendo analizado. Curiosamente, hasta ahora no se han utilizado y hasta donde el autor tiene conocimiento el uso de modelos continuos para reproducir un brote real de sarampión, que es el paradigma de la epidemiología matemática. El modelo SIR, es una herramienta sencilla pero bastante poderosa para hallar los parámetros relevantes del sistema, en este caso, el patrón de contactos caracterizado por la influencia del calendario escolar en Londres, la cantidad de susceptibles y de infectados (representado implícitamente en el ajuste del modelo). Los resultados mostraron consistencia con los estudios previamente realizados e inclusive se capturó un patrón de contactos más acorde al calendario escolar cuando se usaban parámetros regularizadores pequeños (1). La tasa de transmisión era casi idéntica a los trabajos previos cuando se utilizaban parámetros regularizadores altos (mayores a 7) y la cantidad inicial de susceptibles en el año de 1944 hallada en la simulación mostró un valor bastante cercano a una aproximación teórica. En general el trabajo mostró resultados lógicos, acordes a la teoría y a los resultados experimentales obtenidos por otros investigadores.

RECOMENDACIONES

El modelamiento de sistemas siempre se puede mejorar mediante el aumento en la rigurosidad del planteamiento. En los diferentes aspectos de éste trabajo se hicieron diversas suposiciones que no siempre corresponden a un escenario real. Entre los factores que se pueden tener en cuenta para una investigación futura, se encuentra la formulación realista del tiempo de permanencia en los compartimientos, que se describió, pero no se implementó en el problema de optimización. De igual forma, la implementación de factores estocásticos y la variabilidad de las clases epidemiológicas con la edad (no solo con el tiempo) brindaran complejidad matemática a cambio de una mayor precisión en los resultados, reflejados en un mejor ajuste y en una tasa de contactos aún mas precisa.

REFERENCIAS

1. Hethcote, H.W., *The Mathematics of Infectious Diseases*. SIAM Review, 2000. 42(4): p. 599-653.
2. Anderson, R.M., and R.M. May. 1991. *Infectious diseases of humans: dynamics and control*. Oxford University Press, Oxford, UK
3. Hethcote, H.W., *A thousand and one epidemic models*. in Theoretical Biology, S.A.Levin,ed.Lecture notes in Biomath.100, Springer-Verlag, Berlin, 1994, pp.504-515
4. Bjornstad, O.N., B.F. Finkenstadt, and B.T. Grenfell, *Dynamics of measles epidemics: Estimating scaling of Transmission rates using a time series SIR model*. Ecological Monographs, 2002. 72(2): p. 169-184.
5. Liu, W.-m., H.W. Hethcote, and S.A. Levin, *Dynamical behavior of epidemiological models with nonlinear incidence rates*. Journal of Mathematical Biology, 1987. 25(4): p. 359-380.
6. Rohani, P., M.J. Keeling, and B.T. Grenfell, *The interplay between determinism and stochasticity in childhood diseases*. Am Nat, 2002. 159(5): p. 469-81
7. Lloyd, A.L, *Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics*. (2001) Theor. Popul. Biol.60: 59–71.
8. Young, J, Trabajo de investigación realizado en Texas A&M University.

9. Wearing, H.J., P. Rohani, and M.J. Keeling, *Appropriate Models for the Management of Infectious Diseases*. PLoS Medicine, 2005. 2(7): p. e174.
10. Finkenstadt, B.F. and B.T. Grenfell, *Time Series Modelling of Childhood Diseases: A Dynamical Systems Approach*. Applied Statistics, 2000. 49(2): p. 187-205.
11. Cauchemez, S. and N.M. Ferguson, *Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London*. J R Soc Interface. 2008 Aug 6;5(25):885-97
12. Schenzle, D., *An Age-Structured Model of Pre- and Post-Vaccination Measles Transmission*. Math Med Biol, 1984. 1(2): p. 169-191.
13. Grenfell, B.T., O.N. Bjornstad, and B.F. Finkenstadt, *Dynamics of Measles epidemics: Scaling noise, determinism, and predictability with the TSIR model*. Ecological Monographs, 2002. 72(2): p. 185-202.
14. Finkenstadt, B.F., O.N. Bjornstad, and B.T. Grenfell, *A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreaks*. Biostatistics. 2002 Dec;3(4):493-510.
15. Fine, P.E. and J.A. Clarkson, *Measles in England and Wales--I: An analysis of factors underlying seasonal patterns*. Int J Epidemiol, 1982. 11(1): p. 5-14.

ANEXO 1. SOFTWARE EMPLEADO

- **AMPL (versión 2008 09 07)** :(*A Mathematical Programming Language*): Este programa fue el utilizado para programar y resolver el problema de optimización. Este software está diseñado para solucionar problemas de optimización a gran escala. Con el fin de implementar el problema de optimización, fue necesaria la “transformación” de la formulación continua a su versión numérica. Así que la integral del error se convirtió en la sumatoria del error durante el período de estudio. Las ecuaciones diferenciales se plantearon utilizando el método de Euler implícito (con un tamaño de paso lo suficientemente pequeño) y la segunda derivada de la tasa de contactos, mediante diferencias finitas.
- **MATLAB (versión 7.5)**: Con este software fue posible exportar los datos de incidencia del sarampión hacia un archivo que pudiera ser asimilado por AMPL. Posterior a la solución del problema de optimización, se generó un código en AMPL para que creara un archivo *.m* con los resultados obtenidos y así poder visualizar los resultados en MATLAB.
- **DYMOLA (versión 6)**: Es un software de simulación que permite modelar sistemas de una forma mas conveniente que en MATLAB. Con DYMOLA se estudiaron los modelos genéricos y las mejoras hechas a éstos (tiempo de permanencia en los compartimientos)