

ESTIMACIÓN DE PARÁMETROS EN MODELOS
EPIDEMIOLÓGICOS DE VIH/SIDA

PARAMETER ESTIMATION IN HIV/AIDS
EPIDEMIOLOGICAL MODELS*

JORGE BARRIOS GINART[†]

AYMÉE DE LOS ÁNGELES MARRERO SEVERO[‡]

MARTA LOURDES BAGUER DÍAZ-ROMAÑACH[§]

HÉCTOR DE ARAZOZA RODRÍGUEZ[¶]

Received: 11 Nov 2008; Revised: 23 Sep 2009; Accepted: 6 Apr 2010

*Este trabajo ha recibido el apoyo de los proyectos AECID A/010208/07 y D/9842/07.

[†]Facultad de Matemática y Computación, Universidad de La Habana, San Lázaro y L, La Habana CP 10400, Cuba. E-Mail: jbarrios@matcom.uh.cu

[‡]Misma dirección que/*same address as* J. Barrios. E-Mail: aymee@matcom.uh.cu

[§]Misma dirección que/*same address as* J. Barrios. E-Mail: mbaguer@matcom.uh.cu

[¶]Misma dirección que/*same address as* J. Barrios. E-Mail: aragoza@matcom.uh.cu

Resumen

El proceso de validación de modelos matemáticos que describen aplicaciones prácticas implica, generalmente, la estimación de los parámetros desconocidos que en ellos intervienen. En este trabajo, con el fin de estimar estos parámetros en los modelos de VIH/SIDA de [7], se formula el problema de estimación de parámetros en ecuaciones diferenciales ordinarias de primer orden con punto inicial conocido y se presenta una estrategia de solución al mismo. Se verifica, además, cuál es, en algún sentido, el modelo de VIH/SIDA que mejor representa los datos reales existentes según la estrategia de solución.

Palabras clave: Estimación de parámetros, modelos epidemiológicos, búsqueda activa de contactos, mínimos cuadrados ponderados.

Abstract

The validation process of mathematical models that describe practical applications usually implies the estimation of the unknown parameters that are involved. In this work, in order to estimate these parameters in the HIV/AIDS models of [7], the problem of estimating the parameters in first order ordinary differential equations with known start point is formulated and a strategy of solution is presented. It is verified as well, which is the model of HIV/AIDS that represents best the real data according with the strategy of solution.

Keywords: Parameter estimation problem, epidemiologic models, contact tracing, weighted least squares.

Mathematics Subject Classification: 65K05, 65K10, 37N40, 47A10.

1 Introducción

De cara a una aplicación real, la epidemia de VIH/SIDA en Cuba ha sido modelada en [9, 2, 1, 7] como una dinámica de poblaciones, mediante ecuaciones diferenciales ordinarias (**EDOs**).

A diferencia de los sistemas físicos, para los cuales se dispone de un modelo totalmente establecido por leyes bien conocidas, la modelación de procesos epidemiológicos exige una gran dosis de intuición a la hora de establecer las relaciones de causalidad entre las variables. De ahí que, en el estudio de la epidemia VIH/SIDA, a través de los modelos que la describen, es importante el problema de estimar los parámetros que rigen su evolución [3, 6] - para su posterior interpretación -, así como validar la efectividad de cada modelo, o sea, la certeza de las intuiciones utilizadas en su diseño.

En general, el problema de estimación de parámetros en modelos descritos por EDOs, con datos o mediciones reales disponibles en el tiempo,

puede enfocarse como un problema de optimización. En nuestro caso, se minimiza una suma de funciones residuales entre los datos y los valores correspondientes obtenidos de la solución del sistema.

En este artículo se muestra la estrategia de solución que se utilizó para resolver el problema de estimación de parámetros en los modelos de VIH/SIDA considerados. Además, se ordenan dichos modelos - lo que permite seleccionar el que mejor simula o explica los datos reales - según diferentes criterios.

2 VIH/SIDA en Cuba

El Síndrome de Inmunodeficiencia Adquirida (**SIDA**) se define como un conjunto de manifestaciones clínicas que aparecen como consecuencia de la depresión del sistema inmunológico debido a la infección por el Virus de la Inmunodeficiencia Humana (**VIH**).

Los primeros casos de SIDA se reportan al principio de la década de los 80, por lo que es considerada una enfermedad relativamente joven. Sin embargo, se ha expandido geográficamente, y tiene como agravante la inexistencia de un tratamiento inmunizador o cura definitiva, por lo que el proceso infeccioso termina en la muerte del enfermo.

Aunque existe una voluntad global encaminada en la lucha contra el SIDA, según el Programa Conjunto de las Naciones Unidas sobre el VIH/SIDA [13], la región del Caribe clasifica como la segunda más afectada del planeta, sólo superada por África, que sigue siendo el epicentro mundial de la pandemia de SIDA.

En este contexto, Cuba posee la epidemia más limitada de las Américas y una de las más bajas del mundo, teniendo una prevalencia nacional del VIH en adultos inferior al 0,2 por ciento en el 2005 [13].

Un factor decisivo para la baja incidencia de la epidemia en Cuba es, sin dudas, el Programa Nacional sobre VIH/SIDA, establecido en el país desde 1983. Este programa incluye entre sus premisas, la aplicación de medidas de tipo social encaminadas a la educación y la prevención. Un elemento asociado a estas medidas es el Programa de Notificación de Parejas (**PNP**) o búsqueda activa por contactos sexuales, llevado a cabo desde 1986. Mediante el PNP, cuando un individuo es detectado como portador del virus se le sugiere declarar sus contactos sexuales, los que son buscados de manera activa y sometidos a un test de VIH.

El objetivo de la estrategia de búsqueda activa de contactos (**Contact Tracing**¹) se basa en la detección de portadores asintomáticos del VIH

¹Término en inglés muy expandido para la búsqueda activa de contactos.

(seropositivos) mucho antes que desarrollen el SIDA, dada la alta probabilidad que tiene un portador de transmitir la infección por desconocerse seropositivo. Además, una temprana detección permite iniciar un tratamiento que contribuya al retraso de la aparición de los síntomas de la enfermedad y aumente la calidad de vida del infectado.

Para validar el efecto de este método de control en el tamaño de la epidemia, y predecir y/o estimar los parámetros que rigen la evolución de la epidemia se han realizado investigaciones que pueden encontrarse en [3, 6, 9, 2, 1, 7] y otras referencias, las cuales han tenido como eje fundamental la propuesta de modelos descritos por sistemas de EDOs que simulan la detección de la epidemia de VIH/SIDA mediante la búsqueda activa de contactos. El objetivo de estos modelos no es modelar cómo se generan nuevas infecciones por VIH, sino cómo son detectadas las personas infectadas por el virus.

Aunque han sido varios los modelos propuestos para describir la dinámica del VIH/SIDA en Cuba mediante el Contact Tracing, este artículo aborda sólo los presentados en [7], en los que se consideran las siguientes variables de estado:

- $X(t)$: representa el número de personas infectadas con VIH que no conocen que están infectados en el momento t .
- $Y_1(t)$: representa el número de personas infectadas con VIH que conocen que lo están en el momento t y que fueron detectados aleatoriamente (análisis de rutina, prueba a embarazadas, etc).
- $Y_2(t)$: representa el número de personas infectadas con VIH que conocen que lo están en el momento t y que se descubrieron a través de la búsqueda por contactos.
- $Z(t)$: representa el número de personas enfermas con SIDA en el momento t .

La definición epidemiológica de cada uno de los parámetros que intervienen en los modelos se especifican en el Cuadro 1.

Finalmente la dinámica que se describe en [7] mediante el sistema de EDOs es la siguiente:

$$\begin{aligned}
 \frac{\partial X}{\partial t} &= \lambda X + \lambda' (Y_1 + Y_2) - (k_1 + \mu + \beta) X - q(k_2, X, Y_1, Y_2), \\
 \frac{\partial Y_1}{\partial t} &= k_1 X - (\mu + \beta') Y_1, \\
 \frac{\partial Y_2}{\partial t} &= q(k_2, X, Y_1, Y_2) - (\mu + \beta') Y_2, \\
 \frac{\partial Z}{\partial t} &= \beta X + \beta' (Y_1 + Y_2) - \mu' Z
 \end{aligned} \tag{1}$$

Notación	Parámetro
λ	Tasa de contagios causados por individuos seropositivos no detectados.
λ'	Tasa de contagios causados por individuos seropositivos detectados.
k_1	Tasa con que una persona desconocida e infectada con VIH es detectada por el sistema de forma aleatoria.
k_2	Tasa con que una persona desconocida e infectada con VIH es detectada por el sistema a través de la búsqueda de contactos.
β	Tasa con que personas infectadas con VIH, no detectadas, desarrollan el SIDA.
β'	Tasa con que personas infectadas con VIH, detectadas, desarrollan el SIDA (el recíproco del tiempo que va entre Y y Z).
μ	Tasa de mortalidad de la población sexualmente activa.
μ'	Tasa de mortalidad de la población con SIDA.

Cuadro 1: Definición epidemiológica de los parámetros en los modelos.

tal que $X \geq 0, Y_1 \geq 0, Y_2 \geq 0, Z \geq 0$.

Para el término de la búsqueda de contactos $q(k_2, X, Y)$ del sistema (1) se han considerado, en este trabajo, cuatro posibles expresiones, que definen los modelos lineales y no lineales siguientes:

1. El Modelo k_2X (M1) cuando $q(k_2, X, Y) = k_2X$.
2. El Modelo k_2Y (M2) cuando $q(k_2, X, Y) = k_2(Y_1 + Y_2)$.
3. El Modelo k_2XY (M3) cuando $q(k_2, X, Y) = k_2X(Y_1 + Y_2)$.
4. El Modelo $k_2\frac{XY}{X+Y}$ (M4) cuando $q(k_2, X, Y) = k_2\frac{X(Y_1+Y_2)}{X+Y_1+Y_2}$.

Las expresiones consideradas anteriormente para el término $q(k_2, X, Y)$ representan situaciones clásicas en la modelación de dinámicas de poblaciones (ver [12]). En los modelos M1 y M2 el término se plantea mediante expresiones lineales (casos más simples). Luego, en M3 es una forma cuadrática como los que provienen de las dinámicas de gases y por último en M4 se utiliza un término proporcional a la densidad de una de las poblaciones.

Se dispone de datos reales anuales de las variables Y_1, Y_2 y Z durante un período de 20 años desde 1986 hasta el 2005 (o sea el número de personas de cada uno de los grupos por cada año). Naturalmente no es posible obtener

los datos de la variable X por su propio significado en los modelos, sin embargo, se cuenta con una estimación para su valor inicial calculado a partir del número de infectados con VIH descubiertos después de 1986 pero que según los períodos de la infección estarían ya seropositivos en 1986. Esto completa los valores iniciales:

$$X(0) = 230, Y_1(0) = 68, Y_2(0) = 26, Z(0) = 3, \quad (2)$$

(pertenecientes al año 1986 en que se comienza la ejecución del PNP) de las variables de estado para los modelos descritos por el sistema (1).

A partir de estudios anteriores se ha obtenido que los parámetros que intervienen en los modelos se encuentran en los intervalos reflejados en el Cuadro 2. Específicamente los intervalos para β, β', μ, μ' fueron obtenidos por H. de Arazoza con métodos no paramétricos aplicados a los datos reales existentes con intervalos de confianza al 95%. El valor de λ' se calcula como $\lambda' = \lambda\rho$ (ver [7]). Los intervalos para λ, ρ, k_1, k_2 se toman de estudios anteriores realizados por H. de Arazoza et al.

Parámetro	Intervalo
λ	[0.4000, 0.7000]
ρ	[0.0100, 0.1000]
k_1	[0.1000, 0.4000]
k_2	[0.1000, 0.4000]
β	[0.1188, 0.1395]
β'	[0.1538, 0.1739]
μ	[0.0086, 0.0135]
μ'	[0.7042, 0.9259]

Cuadro 2: Intervalos de los parámetros.

3 El problema de estimación de parámetros

En el tratamiento del Problema de Estimación de Parámetros (PEP), una idea muy utilizada se basa en estimar los parámetros desconocidos p_1, \dots, p_k de un modelo matemático que describe una situación de la vida real, mediante la minimización de la distancia entre un conjunto de datos experimentales conocidos y los valores obtenidos del modelo teórico en ciertos puntos. La Estimación de Parámetros (EP) es extremadamente importante en situaciones prácticas como la que se aborda en este artículo — donde se dispone de datos reales y el modelo matemático correspondiente— para analizar el comportamiento del sistema dinámico.

Una formulación general del PEP en EDOs como problema de optimización puede enunciarse de la siguiente manera: Sea el problema continuo de optimización,

$$\begin{aligned} \min_p \quad & \varphi(p) = \sum_{i=1}^s G_i(y(\tau_i, p), \bar{y}_i), \\ \text{s.a.} \quad & \dot{y}(t, p) = f(t, y(t, p), p), \quad t \in [0, T], \\ & y(0) = y_0, \\ & 0 \leq \tau_i < \tau_{i+1} \leq T, \quad i = 1, \dots, s-1, \\ & p_l \leq p \leq p_u, \end{aligned} \tag{3}$$

donde $y_0, y, \bar{y} \in \mathbb{R}^n$; $p, p_l, p_u \in \mathbb{R}^k$; $G_i : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$; $f : [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^n$. Los valores de y_0, p_l y p_u son conocidos.

En (3) se modela un proceso dinámico definido por un sistema n -dimensional de EDOs, que dependen de un vector k -dimensional de parámetros desconocidos p con cotas superiores e inferiores p_l y p_u respectivamente. Además, se consideran un conjunto de datos o mediciones $\{\bar{y}_1, \bar{y}_2, \dots, \bar{y}_s\}$ del vector variable n -dimensional $\bar{y}(\cdot)$ observado en los instantes de tiempo $(\tau_1, \tau_2, \dots, \tau_s)$ respectivamente. En la formulación, las funciones G_i y f se consideran continuamente diferenciables con respecto a sus argumentos. En los modelos considerados en este artículo, el vector de parámetros está representado por $p = (\lambda, \rho, k_1, k_2)$ y los vectores $\bar{y}_i = (\bar{X}(\tau_i), \bar{Y}_1(\tau_i), \bar{Y}_2(\tau_i), \bar{Z}(\tau_i))^T$ denotan las mediciones reales de cada una de las variables de estado del modelo (1) en los tiempos correspondientes a los veinte años en que se dispone de datos.

El objetivo del problema es calcular el valor del vector p que minimice la función φ . Evidentemente, es posible definir más de un criterio de ajuste para el PEP. De manera particular en esta formulación la función φ se expresa como una sumatoria de funciones residuales G_i , las cuales comparan las aproximaciones que devienen de la solución del sistema de EDOs con el conjunto $\{\bar{y}_1, \bar{y}_2, \dots, \bar{y}_s\}$ de datos o mediciones en los correspondientes instantes de tiempo $(\tau_1, \tau_2, \dots, \tau_s)$.

Un criterio a tener en cuenta a la hora de elegir la función de ajuste φ es, sin dudas, la posibilidad de afrontar matemáticamente situaciones particulares del PEP específico que se desea resolver. En el caso de este estudio, no se cuenta con todos los datos necesarios de las variables de estado de los modelos de VIH/SIDA tratados, pues por el propio significado de la variable de estado X en estos modelos se hace imposible tener datos reales de la misma. Por lo tanto, la elección de la función objetivo o de ajuste φ del problema en cuestión debe admitir el tratamiento de situaciones experimentales de este tipo, o sea, donde no todas las componentes del vector

de estado y , pueden ser medidas u observadas; esa es la razón por la cual se apuesta por la formulación clásica de mínimos cuadrados ponderados no lineales, siendo las funciones $G_i(p) = \frac{1}{2}(y(\tau_i, p) - \bar{y}_i)^T W_i (y(\tau_i, p) - \bar{y}_i)$, $i = 1, \dots, s$, donde las matrices W_i son matrices diagonales de ponderaciones con elementos no negativos. Cuando hay ausencia de datos, como se explicó anteriormente, se otorga una ponderación igual a 0 en los elementos diagonales correspondientes a las componentes que no han sido medidas. Por supuesto, es posible elegir pesos distintos de cero acorde a las diferentes precisiones de las componentes del sistema, por ejemplo, sobre la base de alguna información estadística.

El hecho de que la solución del sistema de EDOs no pueda obtenerse analíticamente o que hacerlo desencadenaría un proceso muy complejo, hace que, en general, las restricciones diferenciales deban ser discretizadas de alguna manera, proporcionando un grupo de restricciones algebraicas dependientes del método numérico de integración de EDOs utilizado. Esto convierte el problema continuo en un problema discreto aproximado, como vía de obtener una solución aproximada satisfactoria del PEP en EDOs.

Mediante la aproximación en diferencias de EDOs se llega a la siguiente formulación con restricciones discretizadas

$$\begin{aligned} \min_p \quad & \varphi(p) = \frac{1}{2} \sum_{i=1}^s (y_i(p) - \bar{y}_i)^T W_i (y_i(p) - \bar{y}_i), \\ \text{s.a.} \quad & D_h Y - F(Y, p) = 0, \\ & p_l \leq p \leq p_u, \end{aligned} \tag{4}$$

donde $Y^T = (y_1^T, y_2^T, \dots, y_s^T)$ es la solución numérica de las restricciones diferenciales del problema (3) de acuerdo con una discretización que las convierte en las restricciones algebraicas de (4).

D_h es la matriz correspondiente a la discretización de \dot{y} y F es una función de Y y p , cuya forma depende del método de discretización empleado.

Aunque se han utilizado varios métodos para resolver el PEP en EDOs basados en formulaciones con restricciones algebraicas, que devienen de restricciones diferenciales, en este caso se resuelve el problema sin las restricciones algebraicas a través de la formulación con restricciones sólo para las cotas de los parámetros.

3.1 Formulación con restricciones de cotas

Al obtener en términos de p las variables y_i en las restricciones (4), se llega a la siguiente formulación con restricciones de acotación de tipo caja:

$$\begin{aligned} \min_p \quad & \varphi(p) = \frac{1}{2} \sum_{i=1}^s (y_i(p) - \bar{y}_i)^T W_i (y_i(p) - \bar{y}_i), \\ \text{s.a.} \quad & p_l \leq p \leq p_u, \end{aligned} \quad (5)$$

donde y_i , $i = 1, \dots, s$ es la solución numérica del problema de valores iniciales (3) para un valor determinado del vector de parámetros p en los instantes de tiempo $(\tau_1, \tau_2, \dots, \tau_s)$ de los datos o mediciones.

Una formulación de este tipo para resolver el PEP en EDOs muestra un problema más sencillo de optimización no lineal, donde sólo intervienen restricciones de cotas. Una formulación alternativa para el problema sin restricciones diferenciales es la tipo Lagrange que emplea Marrero en [10]. Sin embargo, para usarla, mediante la versión algorítmica del Teorema Fundamental de [10], se precisa la implementación de un gran número de derivadas, gradientes, etc, para resolver los sistemas de ecuaciones adjuntas o conjugadas que intervienen. Por tanto, esta última no se consideró en este trabajo pues se quiso aprovechar las facilidades de las funciones del software MATLAB para la formulación propuesta.

3.2 Estrategia de solución

La estrategia de solución utilizada para resolver el PEP en los modelos de VIH/SIDA propuestos, se basa en la solución del problema de optimización discretizado con restricciones de cotas. Este problema se resolvió mediante tres vías, una determinista, otra no determinista y una tercera vía híbrida (no determinista + determinista), con el fin de disponer de diferentes alternativas de solución para garantizar resultados más confiables y posibilitar comparaciones.

La minimización de una función que depende de la solución de un sistema de EDOs se compone de dos momentos esenciales:

1. Encontrar la solución del sistema de EDOs (para evaluar la función objetivo φ)
2. Encontrar el valor del vector de parámetros $p = (\lambda, \rho, k_1, k_2)$ que minimice la función φ cuando estos tienen restricciones de cotas.

La disponibilidad de los códigos para realizar estas dos tareas y su idoneidad, depende del sistema operativo que se utilice, así como del

lenguaje o el paquete específico. En este sentido se trabajó sobre el sistema operativo Windows con la ayuda del asistente matemático MATLAB Versión 7.3.0.267(R2006b) de MathWorks, Inc.

Para la integración numérica de los sistemas de EDOs se utilizaron las funciones de MATLAB *ode45*, *ode23*, *ode113*, *ode15s*, *ode23s*, *ode23t*, *ode23tb* [14] y se programaron las funciones *LL* y *LLRK4* que implementan los métodos de Linealización Local [8] y Linealización Local-Runge Kutta de orden 4 [5] respectivamente.

Para la minimización de la función con restricciones de cotas se utilizaron las funciones *fmincon* y *lsqnonlin* del toolbox de optimización de MATLAB [15] que implementan métodos deterministas y la función *ga* del toolbox Algoritmo Genético y Búsqueda Directa de MATLAB [16] que implementa un Algoritmo Genético (método no determinista).

La selección del algoritmo de optimización determina el carácter determinista o no de la solución del PEP, pues elegir el Algoritmo Genético como método de optimización, combinado con cualquier integrador de los mencionados constituye la vía no determinista de la estrategia de solución (una propuesta de este tipo puede consultarse en [11]). Cualquiera de los métodos de optimización propuestos se puede combinar con el uso de cualquier integrador. Si el integrador seleccionado utiliza derivadas de la parte derecha del sistema de EDOs, estas se calculan mediante el uso de las rutinas de diferenciación algorítmica generadas por la herramienta ADiMat versión 0.4-r9 [18].

El método híbrido se constituye a partir de la utilización del Algoritmo Genético, tomando la respuesta del mismo como punto inicial para uno de los algoritmos deterministas implementados en *fmincon* o *lsqnonlin*. Este método tipo híbrido permite refinar la solución óptima que brinda el Algoritmo Genético sobre el supuesto de la hipótesis de proximidad del punto inicial que necesitan los métodos deterministas para su convergencia al óptimo.

4 Experimentos numéricos

Para la realización de los experimentos se confeccionó la Interfaz Gráfica de Usuario PET (Parameter Estimation Tool) sobre MATLAB para resolver el PEP en modelos descritos por EDOs de primer orden con punto inicial conocido como los modelos de VIH/SIDA que se tratan en esta investigación. Con el uso de PET fue posible realizar una vasta experimentación numérica para estimar, con un alto grado de fiabilidad, los parámetros que intervienen en los modelos y a su vez permite elegir el, o los modelos que expliquen o simulen mejor los datos reales existentes. Esta “fiabilidad” está

sustentada en la robustez computacional y estabilidad numérica de las rutinas de MATLAB y la incorporación del cálculo de algunas de las derivadas que intervienen por medio de la Diferenciación Algorítmica, lo que evita los errores de truncamiento propios de la diferenciación numérica.

4.1 Elección de los pesos o ponderaciones

Trabajar con función objetivo de mínimos cuadrados ponderados es esencial en este problema por la carencia de datos de la variable de estado X en los modelos de VIH/SIDA tratados. Sin embargo, tal posibilidad no evita que la variación de los pesos conlleve a un criterio de ajuste diferente, por lo que es imprescindible establecer una conexión entre los valores “más probables” de los parámetros (desde el punto de vista práctico), y los que mejor ajustan los datos reales. Esta conexión se puede establecer definiendo un criterio adecuado para la elección de los pesos. En este estudio, se definieron dos criterios de selección de los pesos involucrados en los experimentos numéricos:

Variante 1: Asignar peso 1 para todos los datos, excepto para los de la variable X de los modelos, que se les asigna peso 0.

Variante 2: Asignar peso $(1/Valor_Del_Dato)^2$ para todos los datos, excepto para los de la variable X de los modelos, que se les asigna peso 0.

Con la elección de los pesos en las Variantes 1 y 2, se pretende dotar a la función objetivo de criterios de ajuste basados en los errores absolutos y relativos, respectivamente, que se cometen en cada una de las aproximaciones. Por un lado, la Variante 2 permite disponer de una estimación del error que no depende de la magnitud de los datos. La naturaleza de la información contenida en los mismos excluye el caso en que los denominadores pudieran hacerse “peligrosamente pequeños”. Esta variante puede resultar muy útil si uno quiere comparar el error residual para las distintas magnitudes de los datos. Por otro lado, al contar con un solo juego de datos reales, la Variante 1 ofrece una idea más “natural” de los residuales para cada uno de los parámetros. Realmente ésta última variante es la que ha aportado en el ajuste, como puede verse en las gráficas 1 y 2, los mejores resultados.

4.2 Mecanismo de calificación de los modelos

Dados los modelos propuestos y para cada uno de ellos el conjunto de parámetros para los cuales se obtiene el mejor ajuste al comportamiento

real de la epidemia, es necesario emplear algún mecanismo de calificación para asignarles un puntaje según cierto criterio con el fin de ordenarlos. Desde esta perspectiva, un criterio que proviene del ajuste de datos y es muy utilizado para juzgar el comportamiento de un modelo es el valor de la función objetivo en el conjunto de parámetros óptimo representando el error así concebido “óptimo”, dado que estamos minimizando. En el caso este estudio, se justifica su uso por el hecho de que la función objetivo depende de los residuales diferencia entre estimaciones y datos. No obstante, al criterio anterior podrían añadirse otros basados en el número de parámetros, el número de datos, etc., pero por la naturaleza de los modelos tratados (donde esas otras características no varían de un modelo a otro), el adicionarlas no aportaría cambios significativos en la calificación. Claro está que al elegir el valor de la función objetivo como criterio de calificación, ante las diferentes variantes de selección de los pesos especificadas anteriormente tendremos funciones objetivo diferentes. Luego, los criterios de comparación estarían sujetos a cada variante de manera independiente.

4.3 Resultados

Después de una extensa experimentación numérica se concluye que el método más efectivo (calculando la efectividad por el valor medio de las funciones objetivos en los óptimos respectivos) fue el método híbrido compuesto por el Algoritmo Genético y la rutina *fmincon*, sin ser (desde el punto de vista numérico) importante la elección del método de integración (si lo fue con respecto al tiempo de cómputo de las simulaciones). Con este método se obtuvieron los resultados que se presentan a continuación, los cuales reflejan la media de cada uno de los parámetros luego de 300 simulaciones.

Se tomó de manera aleatoria en cada simulación los valores para β , β' , μ , μ' en los intervalos reflejados en el Cuadro 2, los cuales permanecen constantes en cada simulación.

En los Cuadros 3 y 4 se presentan las medias para cada uno de los parámetros y la función objetivo en cada modelo, utilizando el criterio de ajuste sujeto a la variante 1 y la variante 2 (ver variantes en la subsección 4.1), respectivamente.

El orden de los modelos según su calificación es:

Con la Variante 1 **M4(mejor) \rightarrow M2 \rightarrow M1 \rightarrow M3**

Con la Variante 2 **M4(mejor) \rightarrow M1 \rightarrow M2 \rightarrow M3**

Independiente del criterio de ajuste, el modelo que mejor simula los datos reales es **M4** y el peor el **M3**, sin embargo, se detecta la influencia

Parámetros	M1	M2	M3	M4
λ	0.60212	0.50323	0.70000	0.51223
ρ	0.10000	0.10000	0.10000	0.10000
k_1	0.26095	0.16476	0.40000	0.18024
k_2	0.10000	0.10000	0.10000	0.15677
F.Obj	339501.72	296208.74	20982606.9681	277775.21

Cuadro 3: Valores medios con los pesos de la variante 1.

Parámetros	M1	M2	M3	M4
λ	0.51198	0.46705	0.40000	0.47472
ρ	0.10000	0.10000	0.10000	0.10000
k_1	0.14670	0.13895	0.40000	0.14665
k_2	0.18475	0.11179	0.40000	0.18475
F.Obj	9.20340	9.30460	32.5625	9.19240

Cuadro 4: Valores medios con los pesos de la variante 2.

del criterio de ajuste en el orden de los modelos en el caso de **M1** y **M2**. Por tanto, se evidencia que uno de los aspectos importantes para una correcta selección de modelos, es un criterio de ajuste idóneo a las necesidades prácticas específicas.

En las figuras 1 y 2 se presentan las gráficas de ajustes del modelo **M4** en los valores medios de los parámetros para cada criterio de ajuste. Como puede observarse, las gráficas correspondientes al criterio de ajuste de la variante 1, simulan mejor los datos reales.

Conclusiones

En este trabajo se formula el problema de estimación de parámetros y se presenta una estrategia de solución que incluye diferentes vías, resultando la vía híbrida la más efectiva. Los valores medios que se obtienen para los parámetros varían de un modelo a otro y dependen del criterio de ajuste utilizado (definido por la elección de los pesos que intervienen en la evaluación de la función de ajuste entre el modelo y los datos).

El mecanismo de calificación para los modelos presentado en este trabajo, generó que el modelo *M4* fuese el que mejor simula o explica los datos reales existentes con el uso de los dos criterios de ajuste utilizados. Según dicha selección, debe considerarse que el vector de parámetros medios calculado para este modelo indica con un mayor grado de veracidad las tasas de la evolución de la epidemia de VIH/SIDA en Cuba. Es importante

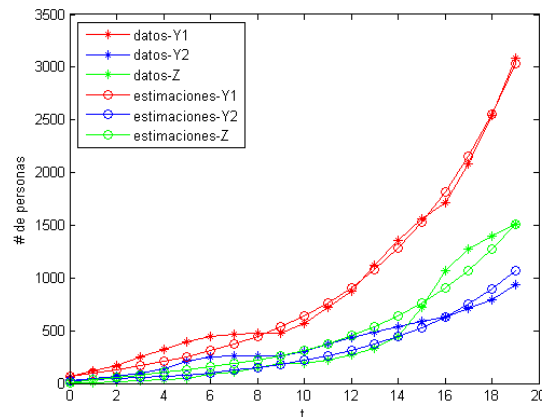


Figura 1: Ajuste del modelo M4 en el vector de parámetros del Cuadro 3 con los datos reales.

señalar que contrario al resultado obtenido en [7] donde se presenta a $M1$ como el mejor modelo, en nuestro caso, según la estrategia que se ha utilizado aquí ese modelo queda en tercer y segundo lugar, respectivamente en cada una de las dos variantes, lo que evidencia la susceptibilidad de resultados de este tipo en relación a los métodos de ajuste y calificación de los modelos, respectivamente.

Sin embargo, el proceso de estimación de parámetros óptimos en cada modelo y de elección del mejor modelo, retroalimenta el proceso de modelación, y en este sentido, cualquier resultado puede no ser definitivo, sino que contribuye a la perfectibilidad de los modelos en aras de alcanzar un comportamiento más cercano a la realidad. Desde esta perspectiva, se expondrán nuevos resultados en futuros trabajos.

Referencias

- [1] Arazoza, H.; Lounes, R. (2002) “A non-linear model for a sexually transmitted disease with contact tracing”, *IMA Journal of Mathematics Applied in Medicine and Biology***19**: 221–234.
- [2] Arazoza, H.; Lounes, R.; Hoang, T.; Interian, Y. (2000) *Journal of Theoretical Medicine* **2**: 267–274.
- [3] Arazoza, H.; Lounes, R.; Pérez, J.; Hoang, T. (2003) “What percentage of the Cuban HIV-AIDS epidemic is known?”, *Revista Cubana de Medicina Tropical*, **55(1)**: 30–37.

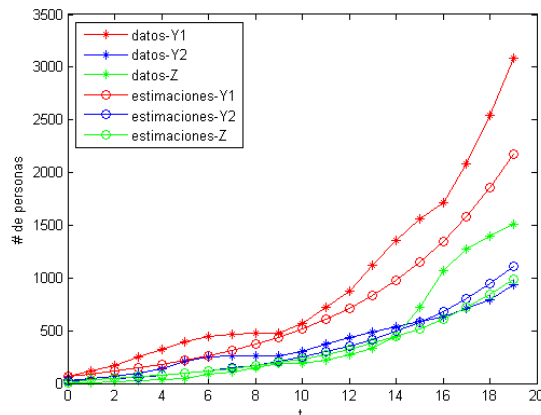


Figura 2: Ajuste del modelo M4 en el vector de parámetros del Cuadro 4 con los datos reales.

- [4] Baden, N.; Villadse, J. (1982) “A family of collocation based methods for parameter estimation in differential equations”, *Chemical Engineering Journal* **23**.
- [5] De la Cruz, H.; Biscay, R.; Carbonell, F.; Jiménez, J.; Ozaki, T. (2006) “Local linearization-Runge Kutta (LLRK) methods for solving ordinary differential equations”, *ICCS, Part I, LNCS 3991*, Springer-Verlag Berlin Heidelberg: 132–139.
- [6] Hsieh, Y.; Arazoza, H.; Lee, S.; Chen, C. (2002) “Estimating the number of Cubans infected sexually by human immunodeficiency virus using contact tracing data”, *International Journal of Epidemiology* **31**: 679–683.
- [7] Hsieh, Y.; Arazoza, H.; Lounes, R.; Joanes, J. (2004) “A class of methods for hiv contact tracing in Cuba: implications for intervention and treatment”, *Deterministic and Stochastic Models for AIDS Epidemics and HIV Infection with Interventions*.
- [8] Jiménez, J.; Biscay, R.; Mora, C.; Rodríguez, L. (2002) “Dynamic properties of the local linearization method for initial value problems”, *Applied Mathematics and Comp.* **126**: 63–80.
- [9] Lounes, R.; Arazoza, H. (1999) “A two type model for the Cuban national programme on HIV/AIDS”, *IMA Journal of Mathematics Applied in Medicine and Biology* **16**: 143–154.

- [10] Marrero, A. (2000) *Un Enfoque para la Solución Numérica del Problema de Estimación de Parámetros en Modelos Definidos por Ecuaciones Diferenciales Ordinarias*. Tesis Doctoral, Universidad de la Habana.
- [11] Marrero, A.; Pedroso, L.; Barrios, J. (2006) “Algoritmos evolutivos en la solución de problemas de estimación de parámetros”, *Revista de Matemática: Teoría y Aplicaciones*, **13**(2): 139–150.
- [12] Murray, J.D. (1993) *Mathematical Biology I. An Introduction*, 2nd edition. Springer, New York.
- [13] Onusida. <http://www.unaids.org>.
- [14] Shampine, L.; Reichelt, M. (1997) “The Matlab ODE Suite”, *SIAM J. Sci. Comput.* **18**(1): 1–22.
- [15] The MathWorks, Inc. *Optimization Toolbox User’s Guide: For Use with MATLAB*. Version 3.
- [16] The MathWorks, Inc. *Genetic Algorithm and Direct Search Toolbox User’s Guide*. Version 2.
- [17] Tjoa, I.; Beigler, L. (1991) “Simultaneous solution and optimization strategies for parameter estimation of differential-algebraic equation systems”, *Industrial Engineering Chemistry* **30**.
- [18] Vehreschild, A. (2004) *The ADiMat Handbook*.