Modelado y análisis de la epidemia VIH-SIDA en Cuba

unia
Universidad Internacional de Andalucía

Premio 2
de estudios Iberoamericanos Grupo La Rábida
Área Científico-Técnica

Modelado y análisis de la epidemia VIH-SIDA en Cuba

Héctor de Arazoza, Aymée Marrero, Elina Miret, Teresita Noriega, Jorge Barrios.
Facultad de Matemática y Computación, Universidad de La Habana, Cuba

EDITA:
UNIVERSIDAD INTERNACIONAL DE ANDALUCÍA
Monasterio de Santa María de las Cuevas
Calle Américo Vespucio, 2
Isla de la Cartuja. 41092 Sevilla
www.unia.es

COORDINACIÓN DE LA EDICIÓN:
Universidad Internacional de Andalucía.

COORDINADOR:

FECHA:
2009

ISBN (soporte papel)
978-84-7993-082-0

Sumario

Palabras Clave

Epidemia VIH-SIDA en Cuba, Modelado de Sistemas, Ecuaciones Diferenciales Ordinarias, Estimación de Parámetros, Análisis Exploratorio de Datos, Técnicas Bio-Inspiradas, Inteligencia Computacional, Algoritmos Genéticos, Redes Neuronales Recurrentes de Hopfield, Mapas Auto-Organizativos de Kohonen.

Resumen

En este trabajo se estudia la epidemia VIH-SIDA en Cuba desde una triple perspectiva: I) se analiza su dinámica de evolución a partir de un modelo de epidemia en Ecuaciones Diferenciales Ordinarias; II) se estiman los principales parámetros de la epidemia, tales como tasas de detección de infectados, tasas de contagio y tiempo de incubación; y III) se lleva a cabo un análisis exploratorio de datos para evaluar la incidencia de la epidemia sobre determinados grupos de la población. Las consecuencias obtenidas están encaminadas a la evaluación y posible mejora de las estrategias seguidas por el sistema sanitario cubano en su tarea de control de la epidemia. Las tareas implicadas en los apartados II) y III) han sido llevadas a cabo mediante técnicas bio-inspiradas (Algoritmos Genéticos) y de Inteligencia Computacional (Redes Neuronales Recurrentes de Hopfield y Mapas Auto-Organizativos de Kohonen).

Esta memoria se presentó para optar por el segundo premio del grupo de Universidades La Rábida en su sección de Ciencias y Tecnología. El jurado decidió otorgarle el premio, con lo cual nos sentimos muy honrados.

Índice de Acrónimos

AG: Algoritmo Genético

BFGS: (Algoritmo de) Broyden-Fletcher-Goldfarb-Shanno

EDO: (Sistema de) Ecuaciones Diferenciales Ordinarias

GN: (Algoritmo de) Gauss-Newton

HSH: Hombres que practican Sexo con Hombres (práctica sexual)

HT: Heterosexual (práctica sexual)

LIP: (Sistema) Lineal en sus Parámetros (Linear In Parameters)

LM: (Algoritmo de) Levenberg-Marquardt

PNP: Programa de Notificación de Parejas (o Búsqueda activa de Contactos Sexuales)

RNRH: Red Neuronal Recurrente de Hopfield

SIDA: Síndrome de Inmunodeficiencia Adquirida

SOM: Mapa Auto-Organizativo de Kohonen (Self Organizing Map)

UNAIDS: Programa de Naciones Unidas para el SIDA

VIH: Virus de Inmunodeficiencia Humana

WHO: Organización Mundial de la Salud (World Health Organization)

Introducción

En este trabajo se aborda el estudio y caracterización de la epidemia de VIH-SIDA en Cuba desde una triple perspectiva:

a. En primer lugar, partimos de un estudio previo sobre un modelo matemático de la epidemia para el que se ha llevado a cabo un análisis de estabilidad y para el que se ha obtenido la expresión del número básico de reproducción (número medio de nuevos infectados que produce una persona infectada a lo largo de su vida) [1]. A partir de este modelo se analiza el estado actual de la epidemia y su tendencia de evolución, se estiman las posibilidades de mantenerla controlada, y se valora el esfuerzo necesario para conseguirlo.

b. En segundo lugar, se lleva a cabo la estimación de aquellos parámetros que intervienen en el modelo de la epidemia que no pueden obtenerse por medios estadísticos. Los parámetros más interesantes para nosotros son las tasas de detección de nuevos infectados por los diferentes programas de búsqueda implementados por el sistema sanitario cubano, las tasas de contagio y la tasa de personas que pasan de seropositivo VIH a enfermo de SIDA sin haber sido detectadas por el sistema sanitario. La determinación de estos coeficientes, proporciona un conocimiento importante sobre el grado de eficiencia de la política de la sanidad cubana en la lucha contra la propagación de la epidemia y proporciona un soporte para la toma de decisiones sobre el mantenimiento o modificación de las diferentes políticas sanitarias desarrolladas.

c. En tercer lugar, se lleva a cabo un análisis exploratorio desde una perspectiva más social de la epidemia, analizando su incidencia para los diferentes grupos poblacionales en función del sexo, prácticas sexuales, edad y nivel de estudios. En este sentido, se ha desarrollado una herramienta que permite clasificar a la población y visualizar de manera gráfica el comportamiento de los diferentes grupos respecto a variables de interés de la epidemia. En concreto, nosotros hemos analizado el tiempo de latencia de la infección (tiempo medio transcurrido desde que una persona es detectada como seropositiva hasta que desarrolla SIDA) para los distintos grupos poblaciones.

Muchas de las operaciones implicadas en las líneas b y c son difícilmente realizables mediante métodos de la matemática y la computación clásica. Así, la estimación de parámetros del modelo matemático no puede ser llevada a cabo mediante los algoritmos clásicos basados en mínimos cuadrados, y el análisis exploratorio de los datos no ha dado buenos resultados utilizando únicamente

métodos de escalamiento multidimensional clásicos. Para soslayar estos inconvenientes se ha recurrido al uso de técnicas de Computación Bio-inspirada y de Inteligencia Computacional tales como Algoritmos Genéticos (AG) y Redes Neuronales Recurrentes tipo Hopfield (RNRH) para la estimación de parámetros, y los Mapas Auto-Organizativos de Kohonen (SOM) para la visualización y análisis exploratorio de datos.

A este respecto, consideramos que este trabajo no sólo representa un avance científico-social en el conocimiento de la epidemia VIH-SIDA, sino también un avance técnico en el conocimiento teórico y de aplicabilidad de las técnicas de Computación Avanzada, materia ésta aún en estado de emergencia para muchos equipos científicos hispanoamericanos.

El resto de esta memoria se desarrolla como sigue: en la sección II, se lleva a cabo una somera descripción de las circunstancias concretas de la epidemia VIH-SIDA en Cuba y se justifica la necesidad de las tareas desarrolladas como apoyo y evaluación de las estrategias sanitarias seguidas para su control, haciendo especial hincapié en la descripción del programa de "búsqueda activa de contactos", que constituye una fuerte apuesta del sistema sanitario cubano en la lucha contra la epidemia; así mismo, se describe el modelo matemático de la epidemia, tratando de hacer compresible el sentido físico -podría decirse social y epidemiológico en este caso- de las ecuaciones y de los parámetros implicados en las mismas. En la sección III, a partir del análisis de estabilidad del modelo previamente realizado, se obtiene un conocimiento matemáticamente sustentado sobre la situación actual de la epidemia, su tendencia de crecimiento, y las posibilidades de mantenerla controlada. En la sección IV, se describe el proceso seguido para la estimación de los parámetros del modelo tanto mediante Algoritmo Genéticos como mediante Redes Neuronales Recurrentes. Una estimación fiable de estos parámetros es esencial para evaluar el grado de efectividad de las estrategias sanitarias establecidas, así como para obtener un valor concreto para las expresiones de los puntos de equilibrio y número básico de reproducción obtenidas en la sección anterior. La sección V está dedicada a la descripción de la metodología que hemos desarrollado para la clasificación y análisis exploratorio de datos, así como su aplicación al análisis de la dependencia del tiempo de latencia con los distintos grupos poblacionales estudiados. La sección VI presenta un resumen de los principales resultados y conclusiones.

Capítulo I.
La epidemia VIH-SIDA en Cuba

A) Una aproximación sanitaria y social.

Las infecciones por VIH y SIDA son dos facetas de un mismo proceso epidémico. Durante la primera, una persona infectada por el virus VIH no presenta síntomas de la enfermedad pero puede transmitirla. Durante la segunda, los síntomas se hacen claramente visibles. Como es sabido, no existe, al día de hoy, tratamiento curativo para el VIH/SIDA, por lo que el proceso infeccioso acaba provocando la muerte del enfermo.

En diciembre de 2000, el informe sobre la epidemia VIH/SIDA del Programa para el SIDA de Naciones Unidas (UNAIDS) y la Organización Mundial de la Salud (WHO) [2], dio una estimación de 36.1 millones de personas vivas infectadas por el virus VIH en el mundo, de las cuales 390,000 vivían en la región del Caribe. Esta región es la segunda en el mundo en porcentaje de prevalencia de la enfermedad después del África Subsahariana. En el informe correspondiente al año 2005, esta organización daba a Cuba una prevalencia del VIH menor del 0.2% en adultos, la más baja en la región del Caribe. Así, la tasa de prevalencia en Barbados y en Republica Dominicana supera el 1%, en Guayana y en Trinidad y Tobago supera el 2%, y excede el 3% en Haití.

Un factor decisivo para esta baja incidencia de la epidemia en Cuba es sin duda el Programa Nacional sobre VIH/SIDA establecido en el país desde 1983. Este programa partía de las siguientes premisas [4]:

I) A través de medios técnico-sanitarios es posible evitar la transmisión por transfusión sanguínea y limitar al mínimo la transmisión perinatal;

II) Las drogas inyectadas no representan un problema en la población, por lo cual su efecto en la transmisión de la epidemia puede ser despreciado; y,

III) La transmisión por vía sexual constituye el principal elemento de riesgo, siendo además un fenómeno imposible de enfrentar exclusivamente con medidas de tipo sanitario.

La premisa I) comportó un conjunto de medidas de control tales como el análisis de todas las donaciones sanguíneas, el análisis de las mujeres embarazadas, el análisis de infectados por otras enfermedades de transmisión sexual y el análisis de enfermos de otras enfermedades como sarcoma o neumonía recurrente. Entre 1986 y 2000 se realizaron al rededor de 23 millones de tests. El resultado es que hasta el año 2000 hubo 10 infecciones por transfusión sanguínea, 2 infectados hemofílicos, y 6 transmisiones verticales (madre a

hijo). La premisa II) parece ser acertada ya que no se ha detectado ningún caso de transmisión asociado al consumo de drogas. Finalmente, la premisa III) comportó la aplicación de medidas de tipo social encaminadas a la educación y la prevención. Asociadas a estas medidas entran en juego dos elementos de este programa que han sido a menudo cuestionados por su posible incidencia en los derechos individuales de las personas: el primero consiste en la instauración de centros sanatorios en los que aquellas personas que daban positivo en un análisis de VIH eran ingresadas, con el objetivo de evitar su participación en la propagación de la epidemia; el segundo, es la puesta en marcha desde 1986 del Programa de Notificación de Parejas (PNP) o de *búsqueda activa de contactos sexuales*, por el cual, cuando un individuo es detectado como portador del virus se le pide que declare cuáles han sido sus contactos sexuales, éstos son buscados de manera activa y sometidos a su vez a un test. El objetivo de esta estrategia de búsqueda activa de contactos (*contact tracing*) es encontrar lo antes posible a aquellos portadores asintomáticos del virus, que aún no han desarrollado el SIDA pero que pueden transmitir la infección entre otras razones porque ignoran su estado. Por otra parte, una temprana detección permitirá un tratamiento paliativo más eficaz y un retraso en la aparición de los síntomas. Como resultado de esta búsqueda activa, el 55% de los detectados con VIH lo son antes de desarrollar el SIDA.

La incidencia del internamiento en sanatorios ha sido drásticamente reducida a lo largo de todos estos años: primero, la gestión de los mismos dejó de estar bajo control policial para quedar en manos de personal sanitario; después, la estancia en estos centros comenzó por reducirse a un tiempo mínimo, permitiéndosele a los internados regresar de manera voluntaria a su lugar de residencia bajo controles sanitarios periódicos; finalmente, los ingresos mismos han sido drásticamente reducidos, sustituyéndose por una asistencia y control ambulatorios llevada a cabo principalmente por médicos de familia especializados.

Sin embargo, a la par que se reducía el factor del internamiento, el Programa de Notificación de Parejas sexuales ha adquirido cada vez más importancia. Esta decisión está avalada por el hecho de que el 90% de los casos de SIDA documentados en Cuba hasta finales de 1997 fueron adquiridos por vía sexual, ya sea Heterosexual (HT), o de Hombres que practican Sexo con otros Hombres (HSH) (preferimos aquí mantener esta notación ya que describe de manera más general un proceso de contagio que no incluye exclusivamente hábitos homosexuales sino también, en un alto porcentaje, hábitos bisexuales). Realmente, la epidemia, que tuvo un marcado carácter HT en sus primeros

años, derivó luego a un carácter principalmente HSH durante los años 90, para finalmente adquirir un carácter mixto en la actualidad. Por otra parte, el más o menos reciente crecimiento del turismo ha originado una emergencia de la prostitución en los últimos años, lo que quizá explique el incremento en el número de positivos VIH a partir de 1996.

Puesto que la puesta en práctica de un programa de búsqueda activa es costoso tanto desde el punto de vista económico como humano, es necesario llevar a cabo una evaluación del mismo para comprobar que sus resultados realmente valen el esfuerzo; de ahí el interés de las autoridades sanitarias cubanas por estimar, con especial atención, el tamaño de la población de infectados por VIH a través de actividad sexual, y la relación de su crecimiento con la estrategia de búsqueda activa planteada. Esta tarea de evaluación ha sido enfrentada, desde una perspectiva matemática, durante más de una década por el **Dpto. de Ecuaciones Diferenciales de la Facultad de Matemática y Computación de la Universidad de La Habana.** Desde el año 2000, este departamento colabora con el **grupo ISIS del Dpto. de Tecnología Electrónica de la Universidad de Málaga** para enfrentar el problema desde la perspectiva de la Inteligencia Computacional. Otras instituciones implicadas en esta investigación son el **Departamento de Tecnología Electrónica de la Universidad de Málaga**, **el Laboratorio MAP 5 de la Universidad Paris Descartes y el Instituto de Medicina Tropical Pedro Kourí de Cuba.**

B) Una aproximación matemática.

La epidemia de VIH/SIDA en Cuba ha sido modelada como una dinámica de poblaciones, mediante Ecuaciones Diferenciales Ordinarias (EDO), con el objetivo de poder obtener una estimación fiable de la población real de infectados por VIH, y de obtener un valor fiable de los parámetros que rigen la evolución de la epidemia [5-6]. Pero a diferencia de otros sistemas mecánicos, para los cuales se dispone de un modelo totalmente establecido por leyes de la Física bien conocidas, el modelado de un “sistema epidémico” exige una gran dosis de intuición a la hora de establecer las relaciones de causalidad entre variables. De ahí que en estos casos, al problema propio de estimar los parámetros que aparecen en el modelo y resolver el sistema de ecuaciones se une el de validar la bondad del propio modelo, es decir, la bondad de las “intuiciones” utilizadas en su elaboración.

El modelo general de la dinámica de la epidemia que nosotros usamos en este trabajo viene dado por el sistema de ecuaciones diferenciales ordinarias (1):

$$\frac{dX}{dt} = (\lambda - k_1 - \beta - \mu)X + \lambda' Y - k_2 f(X,Y) \qquad (1.1)$$

$$\frac{dY2}{dt} = (-\mu - \beta')Y2 + k_2 f(X,Y) \qquad (1.2)$$

$$\frac{dY1}{dt} = (-\mu - \beta')Y1 + k_1 X \qquad (1.3)$$

$$\frac{dZ}{dt} = \beta X + \beta' Y - \mu' Z \qquad (1.4)$$

(1)

Donde Z(t) representa el número de personas con SIDA en el tiempo t.

X(t) representa el número de personas portadoras de VIH que no han sido detectadas en el tiempo t.

Y(t) = $Y_1(t)+Y_2(t)$ representa el número de personas portadoras de VIH que han sido detectadas. De ellas, Y_1 representa el número de portadores detectados por métodos azarosos (donación de sangre, intervención quirúrgica, análisis por embarazo, test anónimo, etc.), mientras que Y_2 representa el número de portadores detectados mediante la estrategia de *búsqueda activa de contactos*.

$f(X,Y)$ representa una función genérica no lineal en X e Y, para la que nosotros consideramos la expresión concreta de $f(X,Y) = \frac{XY}{X+Y}$

Los parámetros implicados en este modelo son los siguientes:

λ: tasa de nuevas infecciones por VIH originadas por la población de seropositivos no detectados X.

λ': tasa de nuevas infecciones por VIH originadas por la población de seropositivos detectados Y.

k_1: tasa de detección de nuevos seropositivos debida a causas azarosas, es decir, sin que intervengan otros seropositivos.

k_2: parámetro indirectamente relacionado con la detección de nuevos seropositivos debida al programa de búsqueda activa de contactos.

β: tasa de personas que pasan de la población de portadores del VIH desconocidos, X, a la población de enfermos de SIDA, Z. Este parámetro es el inverso del tiempo de incubación.

β': tasa de personas que pasan de la población de portadores conocidos, Y, a la población de enfermos de SIDA, Z.

μ: tasa de mortalidad por causas ajenas a la enfermedad del SIDA.

μ': tasa de mortalidad por SIDA.

El esquema de la Figura 1 puede ser útil para comprender la dinámica entre las distintas poblaciones expresada por el modelo sin necesidad de que el lector esté familiarizado con las técnicas de modelado mediante EDOs. Los arcos en línea continua indican transferencia real de individuos entre poblaciones. Los arcos en línea discontinua indican incremento en la población de destino por efecto de contactos sexuales entre un persona portadora de la población de origen y otra sana.

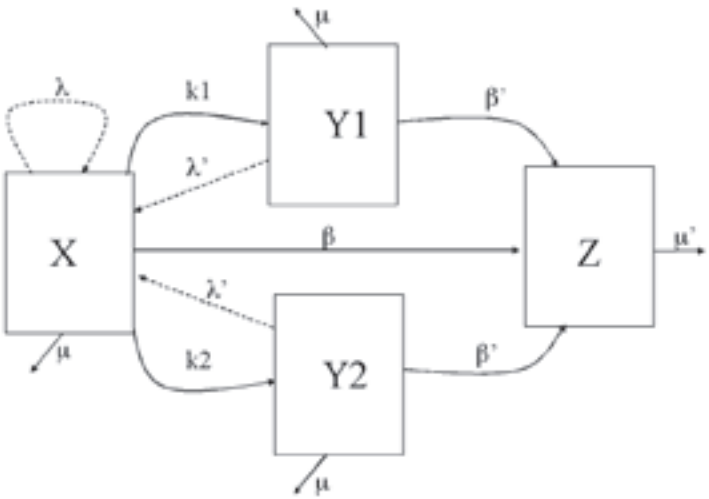


Figura 1. Representación esquemática de la dinámica de poblaciones para la epidemia de VIH/SIDA

Con ayuda del diagrama podemos interpretar las distintas ecuaciones del sistema (1):

La ecuación (1.1) indica que la población de portadores de VIH desconocidos (X) se incrementa en el tiempo con nuevos infectados producidos por la actividad sexual descontrolada de la propia población X (+λX) y de la población Y (+λ'Y), siendo estos incrementos lógicamente mayores cuanto mayores sean las poblaciones X e Y, respectivamente; y siendo λ>λ', ya que, como ha sido comprobado, el comportamiento sexual de las personas que conocen su condición de seropositivos es más responsable que cuando la desconocen, independientemente de su condición social e intelectual. Esta misma población disminuirá debido a la muerte por causas naturales de sus miembros (-μX), disminuirá también por el paso de sus individuos a la población de infectados detectados por azar ($-k_1X$), por el paso de individuos a la población de enfermos

de SIDA ($-\beta X$), y por el paso a la población de infectados detectados por el programa de búsqueda activa ($-k_2 f(X,Y)$). Los tres primeros decrementos serán mayores cuanto mayor sea la población X, mientras que el último depende tanto de X como de Y, ya que son los individuos de esta última población los que deben proporcionar los datos sobre sus contactos.

La ecuación (1.2) indica que la población Y_2 (infectados detectados por búsqueda activa) aumentará en el tiempo por el paso de individuos de X que son detectados como portadores por esa estrategia ($+k_2 f(X,Y)$) (este aumento es justamente igual a la disminución de la población X por esta causa), y disminuirá por efecto de la muerte de sus miembros por causas naturales ($-\mu Y_2$) y por efecto del paso de sus miembros a la condición de enfermos de SIDA ($-\beta Y_2$). Naturalmente, estos dos decrementos será mayor cuanto mayor sea la población Y_2.

La ecuación (1.4) indica que la población de enfermos de SIDA, Z, aumenta por el paso de los infectados por VIH, (conocidos o no), que desarrollan la enfermedad ($+\beta X + \beta' Y$), siendo el incremento mayor cuanto mayores sean estas poblaciones. Esta población Z disminuirá por la muerte de sus miembros ($-\mu' Z$), siendo esta disminución mayor cuanto mayor sea el número de enfermos.

Dejamos al lector, la interpretación de la ecuación (1.3), ya que presenta una gran analogía con la ecuación (1.2).

Entre los parámetros del sistema hay un subconjunto que difícilmente pueden ser estimados a partir de métodos estadísticos, ya que afectan a la población X, que es desconocida por definición. Son éstos los que tienen un mayor interés para nosotros, ya que nos proporcionan una mayor información sobre la epidemia. Estos parámetros son:

k_1 y k_2, ya que nos dan información sobre los tiempos medios de detección de un infectado, permitiéndonos así evaluar la eficacia del programa de búsqueda activa.

β, ya que su inversa nos dan una idea del tiempo medio que un seropositivo tarda en desarrollar la enfermedad o visto de otra forma la proporción de personas que son seropositivas y que acaban desarrollando la enfermedad sin conocer su situación. Este parámetro constituye una autentica piedra de toque de un sistema de prevención de la epidemia, ya que un número alto de β indicaría que hay muchas personas que no pueden ejercer su derecho a una asistencia sanitaria acorde con su situación, y a las que no se les da la oportunidad de llevar a cabo una conducta sexual segura que impida la propagación de la enfermedad.

El parámetro β', tiene menos interés en este trabajo ya que al relacionarse con la población Y puede ser estimado por métodos estadísticos.

λ y λ' son las razones de contagios causados por individuos seropositivos no detectados y detectados respectivamente. Es lógico pensar que λ será mayor que λ', ya que se espera de alguien que conoce su condición de seropositivo una conducta sexual más responsable. El conocimiento experto de la epidemia nos dice que es posible obtener λ' como una proporción r λ, siendo r un valor en torno a 0.069.

Capítulo II.
Estado actual de la epidemia y expectativas de evolución

Una vez que se disponemos de un modelo de la epidemia como el de la ecuación (1), el análisis matemático de sus propiedades nos permitirá responder casi directamente a muchas cuestiones de interés sanitario práctico. Así, lo que se conoce en términos matemáticos como el análisis de puntos fijos o de equilibrio, es decir, la búsqueda de los valores de las variables X, Y_1, Y_2, Z para los cuales el sistema se estabiliza y deja de evolucionar ($\frac{dX}{dt} = \frac{dY}{dt} = \frac{dZ}{dt} = 0$) nos proporciona en términos epidemiológicos el valor de las poblaciones de infectados VIH y enfermos SIDA para los cuales la epidemia permanecería estable. Por ejemplo, el estado (X,Y,Z) = (0,0,0) resulta ser un punto de equilibrio para nuestro modelo que correspondería a una situación de erradicación total de la epidemia. Cualquier otro punto de equilibrio, de existir, representaría una situación de endemia, es decir, una situación en la que la epidemia siempre existe pero no crece ni disminuye.

Para nuestro modelo, el análisis de estabilidad llevado a cabo en [1] nos proporciona un conjunto de puntos de equilibrio que constituye una línea recta de pendiente positiva y paso por el origen en el cuadrante X-Y del espacio de estados. Esta recta es asintótica para todas las trayectorias posibles en el espacio de estados, algunas de ellas con tendencia a infinito y otras con tendencia a cero.

La interpretación de este resultado puede ser la siguiente: los puntos fijos encontrados no son estables, es decir, que la tendencia natural de la epidemia será a de crecer indefinidamente o disminuir hasta desaparecer en función del valor concreto de los parámetros que describen el modelo. Entonces, ¿cual es la situación para nuestro caso concreto?

El parámetro que determina si una epidemia crece o decrece es su Número Básico de Reproducción R_0. Éste puede entenderse como el número medio de infecciones que genera una persona infectada. Si este número es superior a 1, la epidemia crecerá, pero si es inferior a 1 la epidemia decrecerá.

Para el modelo estudiado en este trabajo se ha encontrado la expresión del Número Básico de Reproducción que aparece en la ecuación (2),

$$R_0 = \frac{\lambda}{\mu + \beta + k_1 + \sigma} + \frac{k_1}{\mu + \beta + k_1 + \sigma} \frac{\lambda'}{\mu + \beta'} - \frac{\lambda - \lambda' - (\mu + \beta)}{(\mu + \beta')(\mu + \beta + k_1 + \sigma)} \qquad (2)$$

donde $\sigma = k_2 (x^*)^2$, y x^* es la única solución en el intervalo [0,1] de la ecuación $(\lambda' - \lambda + \beta + k_2 - \beta')\, x^2 + (\lambda - 2\lambda' - k_1 - k_2 - \beta' - \beta)\, x + \lambda' = 0$.

Si sustituimos los valores estimados para los distintos valores de los parámetros que aparecen en la ecuación (2) obtenemos un valor para R_0 en Cuba para la epidemia VIH-SIDA de 1.34, siendo (1.31, 1.36) su intervalo de confianza al 95%. Este valor es muy bajo, especialmente si lo comparamos con el valor que los expertos calculan en general para una enfermedad de transmisión sexual, y que está en el intervalo (2 , 5). En consecuencia, este estudio matemático de la epidemia a partir de su modelado como un sistema dinámico basado en Ecuaciones Diferencias Ordinarias, confirma que la epidemia de VIH-SIDA en Cuba tiene una evolución de crecimiento, aunque por suerte éste es muy lento, mucho menor que en el resto de los países de su entorno, y también mucho menor que el crecimiento esperado para este tipo de epidemia en el mundo.

Sin embargo, las preguntas que surge en este momento de forma natural son: ¿es posible cambiar la tendencia de crecimiento de la epidemia para llegar idealmente a su erradicación?, ¿cual es el esfuerzo en el control de la epidemia que sería necesario por parte del sistema sanitario cubano para alcanzar este objetivo?.

Para dar una respuesta, aunque sea aproximada a estas respuestas, hemos calculado el valor de los parámetros k_1 y k_2 que producirían un valor $R_0<1$ (recordemos que estos parámetros están fuertemente relacionados con la capacidad de detección de nuevos infectados del sistema sanitario cubano). El resultado obtenido es que la epidemia se reduciría, y tendería a erradicarse para valores de $k_1= 0.279$ y $k_2=0.195$.

El valor estimado para el parámetro k_1 por los diferentes métodos reportados en esta memoria queda incluido en el intervalo de confianza de (0.213, 0.224), lo cual supone que como mínimo, habría que conseguir un incremento absoluto de 0.055. Puesto que este parámetro tiene una interpretación directa como el inverso del tiempo de detección por métodos azarosos de una persona infectada por el VIH, este incremento supondría bajar el tiempo de detección en 1 año, es decir, prácticamente en una cuarta parte del tiempo empleado actualmente. Lamentablemente, las posibilidades de obtener una reducción tan drástica del tiempo de detección por medidas azarosas son muy bajas.

Por otra parte, el valor estimado en nuestro trabajo para el parámetro k_2 queda incluido en el intervalo de confianza de (0.181, 0.187). En consecuencia, la consecución de un valor de k_2 que permita el retroceso de la epidemia impondría una disminución en su valor actual del 5% (es necesario advertir que k_2, al no estar linealmente relacionado ni con X ni con Y sino con una expresión no lineal de ambos, no puede ser directamente interpretado como un tiempo de

detección). Esta reducción de k_2 es posiblemente más factible que la necesaria para el parámetro k_1, sin embargo, resultaría enormemente costosa en términos económicos, por lo que será también un objetivo difícilmente alcanzable para un país con la situación económica de Cuba.

En consecuencia, entendemos que debemos tomar conciencia de que la epidemia VIH-SIDA está entre nosotros para quedarse, y que en ausencia de progresos médicos que cambien radicalmente las pautas de su evolución, no nos queda mejor alternativa que tratar de mantenerla en una situación de máximo control posible. En relación con esta idea, la estrategia seguida por el sistema sanitario cubano encaminada a conseguir una detección lo más temprana posible de infectados parece ser en un alto porcentaje, responsable de la muy reducida prevalencia de la enfermedad en relación con el resto de países de su entorno.

Capítulo III.
Estimación de los parámetros del modelo

Una de las tareas más importantes en el estudio de la epidemia VIH-SIDA a partir de su modelado mediante EDOs es la estimación de los parámetros del modelo que no pueden ser aproximados por métodos estadísticos al estar relacionados con la población X, cuyo tamaño es desconocido, al tratarse de personas portadoras asintomáticas de VIH, que no han sido detectadas por el sistema. Entre estos parámetros podemos destacar: k_1 y k_2, coeficientes relacionados con la tasa de detección de infectados por métodos azarosos y por el programa de búsqueda activa de contacto, respectivamente; β, la proporción de personas que pasan por unidad de tiempo directamente de portador VIH desconocido a enfermo sintomático de SIDA (su inversa nos dan una idea del tiempo medio que un seropositivo tarda en desarrollar la enfermedad); y λ y λ', las razones de contagios causados por individuos seropositivos no detectados y detectados respectivamente.

En general, el problema de la estimación de parámetros de un sistema dinámico modelado mediante un sistema EDO, para el que se cuenta con un conjunto de mediciones experimentales a lo largo del tiempo, puede ser enfocado como un problema de optimización, donde la función a minimizar sería una suma de funciones residuales obtenidas de comparar los valores medidos del sistema en los distintos instantes de tiempo con los valores obtenidos de la solución del modelo utilizando un valor estimado de los parámetros. Este proceso queda descrito de manera genérica por la ecuación (3),

$$\begin{aligned} &\min J(\mu) = \sum_{i=1}^{s} G_i\left[x(\mu,\tau_i),\bar{x}^i\right] \\ &s.a. \qquad x'(t) = F[t,x(t),\mu] \quad t\in[0,T] \\ &\qquad\quad x(0) = x^0 \\ &\qquad\quad 0 \le \delta_i < \delta_{i+1} \le T, \quad i = 1,\cdots,s-1 \end{aligned} \tag{3}$$

donde:

$x\in\Re^n$ representa el vector de variables del sistema EDO (en nuestro caso particular el vector de variables (X, Y_1, Y_2, Z));

$\mu\in\Re^r$ representa el vector de los parámetros que deben ser estimados (en nuestro caso, alguna combinación de k_1, k_2, β, λ, λ');

$G_i : \Re^n \to \Re$ es la función residual elegida, calculada para las muestras en cada instante *i* y para un valor estimado particular de los parámetros del sistema en ese instante, y cuya sumatoria para el conjunto de todas las muestra debe ser minimizada.

$x'(t) = F[t,x(t),\mu] \quad t\in[0,T] \quad x(0) = x^0; \quad 0 \le \delta_i < \delta_{i+1} \le T, \quad i = 1,\cdots,s-1$ representa el propio sistema EDO, para cuyas variables se dispone de un valor

inicial *x(0)* y de una serie de muestras tomadas en los instantes τ_i (en nuestro caso, se trata del sistema EDO de la ecuación (1)).

En este trabajo, la función suma de residuos que debe ser minimizada viene dada por la ecuación (4), donde W es definida como una matriz de ponderación.

$$\min_{\mu} \quad \phi(\mu) = \frac{1}{2}\sum_{i=0}^{M}\left(x(t_i,\mu)-\tilde{x}_i\right)^T W_i\left(x(t_i,\mu)-\tilde{x}_i\right] \qquad (4)$$

Para la resolución iterativa de problemas de minimización sin restricciones, se pueden utilizar métodos clásicos que aparecen en la literatura [7-8], y que, por determinadas características, se emplean con frecuencia en la resolución de problemas de estimación de parámetros. Todos tienen como estructura común subyacente, el hecho de que se comienza en un punto inicial, se determina una dirección de movimiento de acuerdo a alguna regla fija y después se sigue esa dirección hacia un punto mejor; en cada nuevo punto se determina una nueva dirección y se repite el proceso hasta encontrar el punto de mínimo relativo de la función objetivo. La diferencia esencial entre ellos radica en la regla mediante la cual se eligen las sucesivas direcciones de movimiento.

Como se mostró en el planteamiento del problema, se trata de la minimización de un funcional sujeto a restricciones diferenciales, el cual puede ser transformado, aplicando técnicas conocidas como el Teorema de Lagrange o más precisamente el Lema Fundamental para el tratamiento de este tipo de problemas, en un problema de minimización sin restricciones, en particular el problema de mínimos cuadrados:

$$min\, s(x) = \frac{1}{2}R(x)^T R(x) = \frac{1}{2}\sum_{j=1}^{m}\left[r_j(x)\right], x\in\Re^n, \text{donde } m > n,\ R:\Re^n \to \Re^m \qquad (5)$$

Los algoritmos más conocidos para la resolución de problemas de mınimos cuadrados no lineales son los de Gauss-Newton (GN) y Levenberg-Marquardt (LM). La diferencia entre ellos es que el primero utiliza la estrategia de búsqueda en la línea, mientras el segundo se basa en región de confianza.

Experimentos numéricos a lo largo del tiempo han mostrado que LM es más estable que GN. Sin embargo, cuando los residuales son pequeños ambos métodos son muy efectivos, lográndose la mayor tasa de convergencia para el caso de residuales cero. Pero, ambos métodos fallan si los residuales son grandes o el problema es altamente no lineal.

La búsqueda de algoritmos efectivos en el caso de residuales grandes, llevó al aprovechamiento de la particular estructura de la función objetivo (suma de cuadrados) y al mejoramiento de métodos quasi-newton entre los que se destaca el llamado Algoritmo BFGS, que logra un mejor desenvolvimiento con residuales grandes, pero que no siempre garantiza las buenas propiedades de GN y LM de ser rápidos en el caso de residuales pequeños.

Todos estos métodos han sido utilizados en la resolución del problema que nos ocupa y aparecen referenciados ampliamente en la literatura, pero todos precisan que todas las funciones que aparecen en el modelo sean continuamente diferenciables. Pero lo más preocupante en el uso de este tipo de métodos clásicos es su eficiencia para problemas de aplicaciones, como el que nos ocupa, para los que se desconocen tanto la magnitud de los residuales, ya que las componentes del vector de parámetros pueden variar en rangos muy diferentes, como la diferenciabilidad de las funciones durante todo el intervalo de definición de las variables. Para evitar estas limitaciones, en este trabajo, se exploran dos vías de solución basadas en métodos no clásicos: los Algoritmos Genéticos (AG), englobados en la categoría de Técnicas Bio-inspiradas; y las Redes Neuronales Recurrentes tipo Hopfield (RNRH), englobadas en la categoría de técnicas de Inteligencia Computacional. Se describe a continuación cada una de estas soluciones.

III.1. Estimación de parámetros mediante Algoritmos Genéticos.

En esta sección describimos de manera detallada los elementos constitutivos de nuestra aplicación de la metodología de Algoritmos Genéticos al problema de la estimación de los parámetros del modelo de epidemia VIH-SIDA de la ecuación (1).

Un Algoritmo Genético (o Evolutivo) AG es un proceso de optimización en el que la obtención de la solución óptima se lleva a cabo mediante la evolución simultánea de un gran número de posibles soluciones en el espacio de búsqueda. Para un problema particular, el AG debe precisar los siguientes componentes [9]:

- Una representación genética para las soluciones factibles del problema. Cada solución representada será llamada en adelante *individuo* o *cromosoma* indistintamente.
- Una forma de crear una población inicial de soluciones factibles o individuos.
- Una *función objetivo* (o función *fitness*) que proporcione información sobre la bondad de cada solución representada y que pueda ser calculada a partir de dicha representación.

• Los *operadores genéticos* que alteren la composición de los individuos de una población, generando así nuevos individuos o soluciones.

• La estrategia de *selección* de individuos de una generación a la siguiente.

• Algoritmo general que determina el sentido de evolución de los individuos y el modo de interacción del resto de los componentes descritos.

• Un criterio de parada del algoritmo.

Describimos aquí cada uno de los elementos anteriores y su caracterización para el problema particular de la estimación de parámetros que nos ocupa.

Representación de una solución factible (definición de un individuo)

Cada individuo viene descrito por un vector $v = (v_1, \ldots, v_p)$, donde cada v_i representa el valor de uno de los p parámetros del modelo que deben ser estimados. Cada componente v_i podrá tomar un valor real en un intervalo particular $[\alpha_i , \beta_i]$ previamente suministrado por un especialista en la epidemia, y que permite centrar la búsqueda en un rango concreto de sus posibles valores.

Se han llevado a cabo distintas implementaciones del algoritmo genético utilizando diferentes vectores de parámetros a estimar. En esta memoria reportamos los resultados obtenidos para la prueba que nos parece de mayor interés, la estimación de los parámetros (k_1, k_2, β, λ).

Generación de una población inicial

La población inicial se crea de manera aleatoria con una distribución uniforme en el conjunto $[\alpha_1, \beta_1]\times [\alpha_2 , \beta_2]\times\ldots\times [\alpha_p , \beta_p]$ donde los parámetros están definidos. En general, en caso de que se conociera alguna información adicional sobre los parámetros, se tomaría en cuenta para crear la población. El número N+1[1] de individuos de una población puede ser libremente establecido, aunque en nuestro caso debemos considerar un compromiso entre la necesaria variedad genética de la población (un número N alto) y el coste computacional que supondrá el cálculo de la función objetivo para cada individuo (un número N bajo). Los resultados presentados en esta memoria se obtuvieron para N=100.

(1) La razón para indicar N+1 como número de individuos estriba en que al aplicar el criterio elitista, en cada momento habrá un individuo destacado (individuo élite) obtenido mediante réplica del mejor individuo de ese momento, el cual no participará en ninguno de los procesos evolutivos de reproducción y mutación a los que es sometido el resto de la población.

Cálculo de la función objetivo para cada individuo

La función objetivo utilizada es el valor cuadrático del error de predicción de la epidemia (ver ecuación (4)). Es decir, para cada individuo de la población genética se debe resolver el sistema EDO que modela la epidemia (ecuación (1)), obteniendo así los valores predichos de las poblaciones X, Y1, Y2, Z en cada instante de integración; a continuación se obtiene la suma de los errores entre los valores predichos y los medidos a lo largo de los distintos instantes de integración para cada una de las poblaciones X, Y1, Y2, Z.

Nuestro equipo de trabajo ha desarrollado una herramienta de simulación llamada **GAPEST** que permite la implementación de diferentes métodos de integración del sistema EDO, entre ellos, las formulas de Dormand-Prince de órdenes 4-5 (caso particular de Runge-Kutta con v estados) y el método de Rosenbrock. En este trabajo, el método de integración elegido ha sido el Runge-Kutta 4-5.

Es necesario resaltar que ésta es la operación de mayor coste computacional del algoritmo, ya que supone la solución del sistema EDO para cada uno de los N individuos de la población Genética.

Descripción de los operadores genéticos utilizados

En nuestro trabajo hemos usado dos operadores genéticos destinados a generar nuevos individuos: *cruce y mutación*. Aunque la herramienta **GAPEST** permite la implementación de diferentes modalidades de estos operadores, los finalmente reportados en esta memoria son el operador *cruce lineal* y el *operador mutación* aleatoria.

En la operación de cruce lineal se eligen al azar dos individuos de la población actual EMBED $v = (v_1, \cdots, v_p)$ y $w = (w_1, \cdots, w_p)$, se generan tres nuevos individuos $0.5(v+w)$, $1.5v - 0.5w$ y $-0.5v + 1.5w$, y finalmente se toman como descendencias los dos con menor valor de la función objetivo. En este proceso, cada individuo es elegido con una probabilidad p_c. Esta operación se realiza N/2 veces. En este trabajo se ha tomado un valor $p_c = 0.75$.

La operación de mutación se aplica a cada uno de los componentes de todos los individuos de la población. Cada componente será mutado con una probabilidad predeterminada p_m. El nuevo valor de componente mutado v_i será un valor aleatoriamente obtenido de su intervalo de definición $[\alpha_i, \beta_i]$. El operador de mutación tiene como principal objetivo sacar al algoritmo de genético de un posible mínimo local en el espacio de búsqueda mediante un

cambio aleatorio en los individuos de la población actual, de ahí que su valor no debe ser excesivamente alto, lo que convertiría al algoritmo en un proceso prácticamente aleatorio. En este trabajo se ha tomado $p_m = 0.01$.

Es necesario hacer notar que los procesos de cruce y mutación vienen restringidos por la aplicación de *elitismo*. Esto es, una vez evaluado cada individuo de la población, se selecciona aquel que representa un menor valor de la función objetivo (el que representa a la mejor solución de la población genética actual). Este individuo es replicado, y su copia es separada de la población de manera que no será sometida a los operadores y pasará directamente a la siguiente generación.

Estrategia de selección de individuos de una generación a la siguiente

La estrategia de selección finalmente llevada a cabo en este trabajo ha sido la de *torneo*. En este caso se comparan dos individuos de la población elegidos al azar, y de ellos se selecciona el que presente un menor valor de la función objetivo (mejor solución). Cada individuo compite exactamente dos veces, lo que hace posible completar una población intermedia de hijos de igual tamaño que la de los padres. Este operador garantiza dos copias del mejor individuo y la desaparición del peor en la siguiente generación.

Algoritmo de evolución

Finalmente, la versión concreta del algoritmo de evolución implementada responde al siguiente pseudo-código:

t ← 0
Generar *P(0)* (población inicial),
Generar EMBED Equation.3 y ubicarlo como *elemento élite* en *P(0),*
Para *t=0,1,2,…,UltimaGeneracion-1,*

1. Hacer cruces entre individuos de *P(t)*, excepto el *elemento élite,*
2. Hacer mutaciones a los individuos de *P(t)*, excepto el *elemento élite,*
3. Evaluar *f* en los individuos de *P(t),*
4. Seleccionar el nuevo *elemento élite* EMBED Equation.3 ,
5. Replicar EMBED Equation.3 y ubicarlo como *elemento élite* de *P(t+1),*
6. Seleccionar de *P(t)* los *N* individuos restantes para completar la población *P(t+1).*

Nótese que, para evitar una operación de evaluación innecesaria de la población inicial, $x^{(0)}$ es generado aleatoriamente, de manera que no tiene por qué ser el mejor individuo en *P(0)*.

Como puede deducirse del pseudo-código, el *criterio de parada* utilizado es el de llevar a cabo un número predeterminado de iteraciones del algoritmo. En este trabajo el número total de iteraciones se ha tomado como *UltimaGeneracion* = 200.

Se han llevado a cabo 100 ejecuciones diferentes del algoritmo, siendo los valores óptimos de los parámetros los que aparecen en la tabla 1.

Tabla 1.- Parámetros estimados mediante la aplicación del Algoritmo Genético y su desviación típica

PARAMETERO	MEDIA	SD
Λ	0.57475	0,027602
k1	0,218515	0,02916
k2	0,184758	0,015654
B	0,129768	0,005808

La gráfica de la figura 2 permite la comparación entre los valores predichos a partir de los parámetros estimados y los valores medidos para Y1 (azul), Y2 (naranja) y Z (magenta). Podemos concluir que los valores estimados son satisfactorios, teniendo en cuenta que por imposición del Algoritmo Genético, éstos deben considerarse como valores constantes a lo largo del tiempo, por tanto no pueden describir las posibles fluctuaciones que pueda sufrir en el sistema real en distintos momentos.

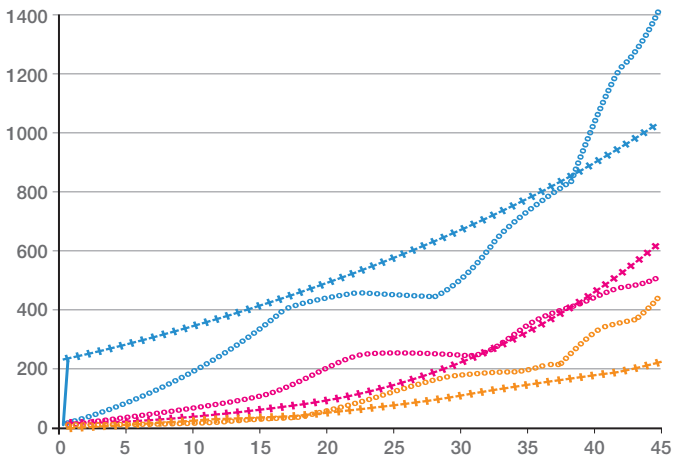


Figura 2.- Curvas medidas (o) y predichas (+) de evolución de las poblaciones Y1 (azul y superior), Y2 (naranja e inferior) y Z (magenta e intermedia).

Algunos comentarios sobre el método basado en Algoritmos Genéticos y sobre los resultados obtenidos.

Los valores obtenidos para los parámetros están cercanos a los que se pueden estimar a partir de otros modelos o con otros métodos, así β es el inverso del periodo de incubación, y nos da un valor medio de 7,7 años. Este periodo se ha estimado por métodos no paramétricos en alrededor de 8 años.

Un objetivo esencial en epidemiología es calcular el Número Básico de Reproducción de la epidemia (R_0) que se define como el número de nuevas infecciones que genera cada persona infectada. Si R_0 es menor que 1, la epidemia decrece en tamaño con el tiempo, si es mayor que 1, la epidemia crece, tanto más rápidamente cuanto mayor sea el número. Con los valores que hemos estimado de los parámetros tenemos un R_0 de 1.34, con un intervalo de confianza (1.31 , 1.36). Es decir que los valores encontrados dan una epidemia no muy fuerte, lo cual concuerda con la situación actual de Cuba.

En cuanto al método computacional utilizado, por su carácter heurístico, podemos afirmar que brinda enormes posibilidades en el tratamiento de este tipo de problemas, particularmente cuando surgen de modelar fenómenos de las ciencias aplicadas, en las que muchas veces los métodos tradicionales no se comportan eficientemente.

Aunque en la salida de la herramienta **GAPEST** sólo se muestra, a modo de ilustración, los resultados para una de las tantas simulaciones obtenidas, hemos trabajado con vectores de parámetros de diferentes dimensiones y hemos obtenido resultados que están en el rango de lo factible o esperado. Este software experimental y en perfeccionamiento, permite diseñar nuevos operadores genéticos y trabajar con los parámetros de estos, con vistas a conservar y mejorar los mejores individuos y no necesita el cálculo de derivadas.

Sin embargo, grandes poblaciones producen muchas evaluaciones de la función objetivo ya que se necesita el fitness de cada individuo y aunque en la mayoría de los casos, los valores óptimos del vector de parámetros son los esperados o deseados, no siempre se dispone de resultados teóricos sobre la convergencia del algoritmo genético. Como ha sido indicado, la evaluación de cada individuo en una población genética dada requiere la integración del sistema EDO para los valores de los parámetros proporcionados por ese individuo. Esta operación constituye la verdadera piedra de toque de nuestro algoritmo genético, ya que siendo de un elevado coste computacional, debe realizarse para cada uno de los individuos de cada una de las generaciones,

es decir ($N \cdot UltimGeneracion$) veces ($2x10^4$ veces en nuestros experimentos). Por otra parte, la convergencia de los métodos de integración utilizados no está garantizada para un sistema EDO.

El problema del coste computacional del algoritmo no supone una traba insalvable para su aplicación siempre que contemos con un computador de la suficiente potencia (cosa totalmente factible hoy en día), ya que la identificación de la epidemia VIH-SIDA que nos ocupa no es, obviamente, un sistema que deba resolverse en tiempo real. Sin embargo, el problema de la convergencia de los métodos de integración si debe ser considerado con atención, y en ese sentido deben explorarse nuevos métodos.

Por otra parte, la definición de los individuos de la población genética lleva implícita la consideración de que los parámetros que deben ser estimados son constantes a lo largo del tiempo de integración del sistema EDO. Esta restricción no tiene por qué ser necesariamente cierta en nuestro caso de estudio, ya que las diferentes tasas de detección de infectados, las tasas de infección o el tiempo de latencia pueden variar en el tiempo como consecuencia de un incremento en la efectividad del programa de búsqueda activa, en el aumento de los tests anónimos, en la emergencia de nuevos mecanismos de detección como puede ser el basado en el médico de familia, en la mejora de la seguridad en la práctica de sexo en la población, o en el descubrimiento de nuevos fármacos entre otra causas. Sin embargo, en el periodo de tiempo estudiado, las variaciones de los parámetros no se consideran lo suficientemente importantes como para que el valor obtenido mediante nuestro algoritmo no pueda ser considerado como un valor medio de utilidad.

IIII.2. Estimación de parámetros mediante Redes Neuronales Recurrentes tipo Hopfield.

En esta sección presentamos un nuevo método para la estimación de los parámetros del modelo de epidemia que nos ocupa, que trata de soslayar las principales limitaciones del método basado en Algoritmos Genéticos anteriormente expuesto, a saber: a) necesidad de integrar el sistema dinámico para cada individuo de cada generación (alto coste computacional y posibilidades de no convergencia) y b) necesidad de considerar los parámetros constantes en el tiempo.

El nuevo método consiste en el diseño de una variante de modelo neuronal de Hopfield que en líneas generales atiende a la siguiente estrategia [10,11]:

Dado un sistema dinámico descrito por un sistema EDO lineal en sus parámetros (p. ej., nuestro modelo de epidemia de la ecuación (1)), se obtiene la expresión del error de predicción del modelo para un determino valor estimado de sus parámetros. A continuación se construye una Red Neuronal Recurrente de Hopfield (RNRH) tal que tenga como función de energía o función de Lyapunov esta función de error de predicción, y cuyos estados (vector de salidas de sus neuronas) representen en cada instante el valor actual del vector de parámetros a estimar. Por su propia naturaleza, esta red de Hopfield evolucionará de manera natural hasta alcanzar un mínimo de su función de energía, en consecuencia, el valor de salida de sus neuronas en su estado estable representará la mejor estimación del valor de los parámetros del sistema dinámico, es decir, aquellos valores de los parámetros que hagan mínimo el error de predicción.

Es necesario destacar que la red de Hopfield no está modelando al sistema dinámico, sino que es un sistema totalmente diferente cuya función energía coincide con el error de predicción del sistema dinámico modelado. Así, la ventaja principal del método es que la integración en el tiempo del sistema que representa a la red de Hopfield es mucho más sencilla y rápida que la integración del sistema dinámico que se quiere identificar, y la integración de éste último no tiene que ser llevada a cabo en ningún momento del proceso.

Otra característica específica de esta variante del modelo de Hopfiled es la variablidad de sus pesos a lo largo de la evolución. Esta variabilidad hace necesaria una nueva prueba de estabilidad para la red, ya que las hasta ahora realizadas presuponían pesos constantes en el tiempo. Este nuevo estudio de estabilidad puede encontrarse en [12].

Describimos a continuación, en detalle, el proceso de implementación de nuestra red de Hopfield:

Como ha sido dicho, las RNRH son sistemas dinámicos cuya estabilidad viene definida por una función de Lyapunov. Dichos sistemas, por tanto, viene caracterizados por tres elementos: sus ecuaciones dinámicas, la función de activación de sus neuronas y la expresión de su función de Liapunov o función de energía. En el caso particular de la formulación de Abe [13], estos elementos vienen definidos, por la ecuación (6).

$$\frac{du_i}{dt} = \sum_j w_{ij} s_j - I_i \qquad \text{6.1.}$$

$$s_i = g(u_i / \gamma) \qquad \text{6.2.}$$

$$E = -\frac{1}{2}\sum_i \sum_j w_{ij} s_i s_j + \sum_i I_i s_i \qquad \text{6.3.}$$

(6)

donde s_i representa la salida de la i-ésima neurona, w_i y I_i son, respectivamente, los pesos y bias de la red, y γ es un parámetro que permite controlar la pendiente de la función de activación de las neuronas (en este caso la función tangente hiperbólica).

El proceso de aplicación de las redes de Hopfield a la optimización de una función objetivo [11] consiste en identificar esta función con la función de energía de la red. A partir de esta identificación, la salida de cada neurona será asociada a una variable de la función objetivo, y los parámetros de dicha función objetivo permitirán obtener los pesos y bias de la red. Esta red así construida evolucionará hasta estabilizarse en un estado de mínima energía, que coincidirá con un mínimo de la función objetivo.

Esta estrategia de optimización puede ser aplicada a la estimación de parámetros de un sistema dinámico como el que nos ocupa en este trabajo, mediante la siguiente metodología:

Partimos de un modelo del sistema dinámico descrito mediante un sistema EDO que suponemos lineal en sus parámetros (LIP), aunque no necesariamente lineal en sus variables de estado, (p.ej., el modelo de la epidemia de la ec. (1), el cual es lineal en sus parámetros, pero no lo es en sus variables ya que nos aparece el término XY/X+Y). Consideramos que algunos de estos parámetros son desconocidos o inciertos (p. ej., los parámetros k_1, k_2, y β). Puesto que el el sistema es LIP puede ser escrito como en la ecuación (7),

$$\frac{dx}{dt} = A(x(t))\theta(t) + b(x(t)) \qquad (7)$$

Donde θ es el vector de parámetros y *A* es una matriz cuyos elementos vienen dados por expresiones no necesariamente lineales de las variables de estado.

Esta notación puede ser simplificada si definimos el vector $y = \frac{dx}{dt} - b$, de manera que el sistema EDO pueda ser reescrito como

$$y = A\theta \qquad (8)$$

Aceptamos la suposición de partida de que el mejor vector de parámetros será aquel que genere unas soluciones del modelo más acordes con las salidas realmente medidas del sistema real. En consecuencia, podemos transformar nuestro problema de estimación paramétrica en otro de optimización, mediante la búsqueda de los parámetros del sistema que minimicen la función de error de predicción en cada instante $e = y - A\hat{\theta} = A\tilde{\theta}$, donde *y* es el valor real de la salida del sistema, $\hat{\theta}$ representan el vector de parámetros estimados en cada instante, y $\tilde{\theta} = \theta - \hat{\theta}$ representa la diferencia entre el valor real y estimado del vector de parámetros.

La función objetivo queda en nuestro caso como

$$V = \frac{1}{2}\|e\|^2 = \frac{1}{2}e^T e = \frac{1}{2}(A\tilde{\theta})^T(A\tilde{\theta}) = \frac{1}{2}\tilde{\theta}^T A^T A\tilde{\theta} \qquad (9)$$

Si, como ha sido indicado previamente, identificamos la ecuación (9) con la ecuación (6.3) obtenemos la expresión de los pesos y *bias* de nuestra red neuronal como los dados por la ecuación (10)

$$W = -A^T A; \qquad I = -A^T y \qquad (10)$$

Un análisis detallado de esta metodología puede ser encontrado en [13].
Como ya ha sido dicho, el modelo de la ecuación (1), en que centramos este trabajo, contiene distintos parámetros que pueden ser considerados como la razón de transmisión entre las respectivas poblaciones. Algunos de estos parámetros son fácilmente medibles o al menos, deducidos de manera fiable por métodos estadísticos. Sin embargo, otros parámetros son verdaderamente difíciles de obtener, como la tasa de detectados VIH mediante el programa de búsqueda activa (k_2), la tasa de detectados mediante métodos azarosos (k_2), o la tasa de transición directa de infectado VIH desconocido a enfermo de SIDA (β). La estimación de estos tres últimos parámetros será nuestro objetivo en este apartado.

Los datos sobre la epidemia VIH-SIDA en Cuba están disponibles con una periodicidad semanal. Sin embargo, se encuentran algunos datos erróneos debido, especialmente, a retrasos en su registro. Para soslayar este problema, nosotros modificamos la base de datos, calculando la cantidad de individuos en cada población con una periodicidad de 4 meses. Con esta periodicidad disponemos de las poblaciones Y1, Y2 y Z, pero la población X es por definición desconocida. Sin embargo, la proporción de X comparada con Y puede ser estimada como X=0.3Y, con ajustes adicionales proporcionados por los expertos sanitarios.

Una vez determinado el vector de parámetros a ser estimados $\theta = (k_1, k_2, \beta)$, la red neuronal puede ser construida a partir de la ecuación (10). Lógicamente, nuestra red neuronal tendrá tres neuronas, una por cada uno de los parámetros a estimar.

Sin embargo, la simulación de la evolución de la red presenta un problema originado por la limitación en la frecuencia de los datos. En este caso, no podemos considerar al sistema como continuo, sino como discreto. En consecuencia, la formulación de su evolución vendrá dada por (11).

$$x_{n+1} = x_n + \Delta t\, A(x_n)\theta + y \qquad (11)$$

donde $\Delta t = \frac{1}{3}$ corresponde al periodo de disponibilidad de los datos (una tercera parte del año).

La formulación de la ecuación (11) nos obliga a reescribir la fórmula del error de predicción en el instante t+1 para un determinado vector estimado $\hat{\theta}$, como

$$e_{n+1} = x_{n+1} - x_n - \Delta t\, A(x_n)\hat{\theta} - y \qquad (12)$$

y la función objetivo se corresponde con la expresión $V = \frac{1}{2} e^T e$, de donde, después de algunos cálculos, la expresión de los peso y *bias* de la resulta:

$$\begin{aligned} W &= -\Delta t^2 A^T A; \\ I &= -\Delta t\; A^T (x_{n+1} - x_{n+1} - y) \end{aligned} \qquad (13)$$

Los resultados de simulación para nuestra red se muestran en las figura 3, 4, 5, 6 y 7 referidas respectivamente, a la representación de los parámetros estimados y a las curvas de predicción de la evolución de las diferentes poblaciones.

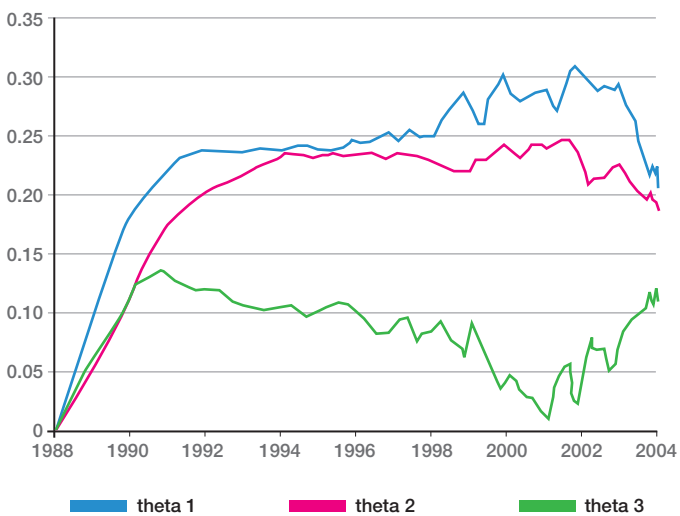


Figura 3.- Representación gráfica de los valores k_1 (theta1 en la leyenda de la gráfica), k_2 (theta2) y β (theta3).

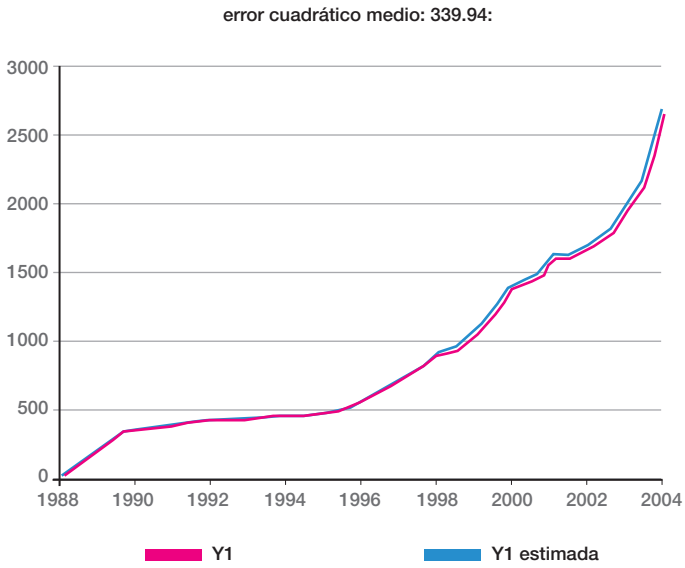


Figura 4.- Representación gráfica de la evolución temporal de Y1: en rojo datos medidos y en azul datos predichos por el modelo para los parámetros estimados mediante un estimador neuronal.

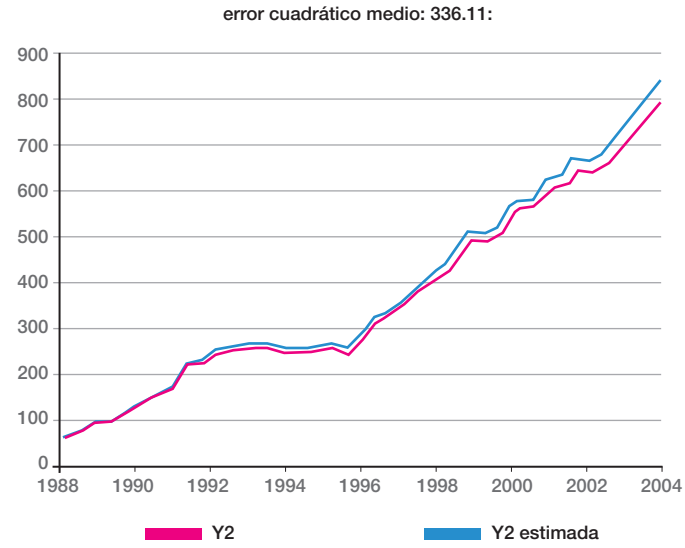


Figura 5.- Representación gráfica de la evolución temporal de Y2: en rojo datos medidos y en azul datos predichos por el modelo para los parámetros estimados mediante un estimador neuronal.

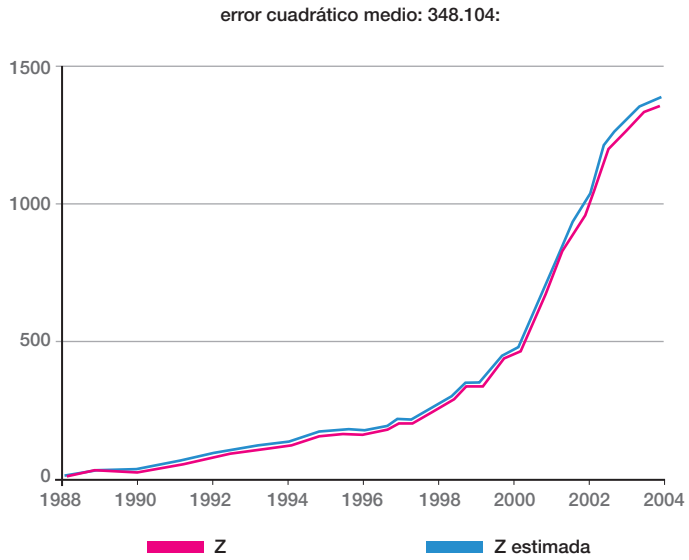


Figura 6.- Representación gráfica de la evolución temporal de Z: en rojo datos medidos y en azul datos predichos por el modelo para los parámetros estimados mediante un estimador neuronal.

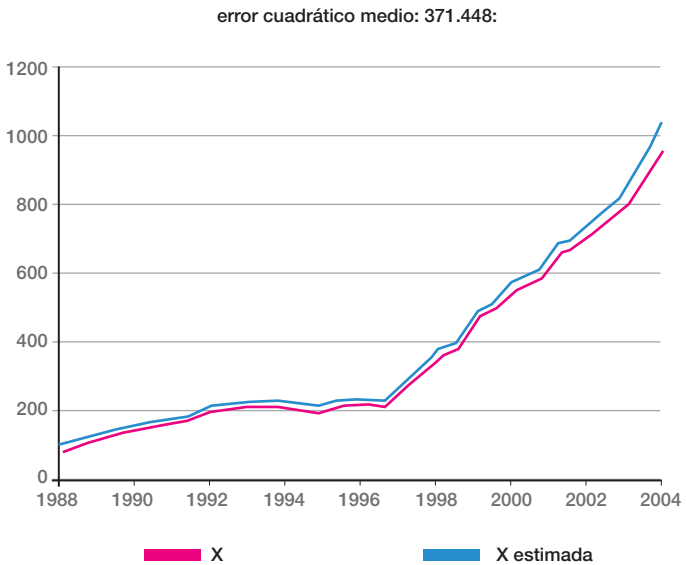


Figura 7.- Representación gráfica de la evolución temporal de X: en rojo datos estimados a partir de Y y en azul datos predichos por el modelo para los parámetros estimados mediante un estimador neuronal.

Algunos comentarios a los resultados obtenidos

El ajuste entre las curvas predichas y reales parece suficientemente bueno como para afirmar que el objetivo de estimación de parámetros ha sido conseguido satisfactoriamente. Por otra parte, los resultados de estimación obtenidos son plausibles y coherentes con las consideraciones de los expertos sanitarios. En particular, el valor obtenido del parámetro β, que oscila entre (0.11 y 0.13) coincide de manera muy aproximada con el obtenido por otros métodos: β es el inverso del tiempo medio que una persona permanece en X antes de desarrollar los síntomas de SIDA. Este tiempo de incubación resulta ser de entre 7.7 y 9 años según nuestros cálculos, lo que coincide con el valor obtenido por medio de Algoritmos Genéticos (7.7 años) y con el valor estimado por los expertos para esta enfermedad (8 años). La oscilación de los resultados después del año 2000 sugiere un cambio en la composición de las distintas poblaciones, lo que conlleva la necesidad de desarrollar modelos más sofisticados que incluyan, por ejemplo, el papel de factores de detección emergentes como puede ser el basado en los médicos de familia.

Finalmente, la variabilidad mostrada por los parámetros a lo largo de los años de estudio de la epidemia justifica el uso de esta metodología neuronal frente a los Algoritmos Genéticos, que como fue dicho en su momento, exigen la suposición de parámetros constantes.

Los resultados de predicción y la bondad de los parámetros estimados paracen también confirmar que el modelo de epidemia propuesto se ajusta suficientemente a la dinámica real.

En el aspecto de investigación computacional, la puesta en práctica de este método de estimación de parámetros nos ha permitido poner a prueba el estimador neuronal implementado respecto a su capacidad de trabajar con un sistema no continuo, en que no podemos hablar de excitación exterior.

Capítulo IV.
Análisis exploratorio de datos en la epidemia VIH-SIDA en Cuba mediante Mapas Auto-Organizativos de Kohonen

En esta última línea de trabajo, abandonamos el enfoque analítico de la epidemia VIH-SIDA para sustituirlo por otro de carácter más social. Nos planteamos la cuestión de cómo afecta la epidemia a las personas en función del grupo poblacional al que pertenecen, es decir, pretendemos contestar a cuestiones como ¿afecta la epidemia VIH-SIDA de la misma manera a hombres que a mujeres?, y dentro de cada sexo ¿afecta la epidemia de la misma manera a todos los intervalos de edad?; ¿tiene la epidemia una incidencia distinta para los diferentes grupos profesionales o culturales?.

Este enfoque presenta un conjunto de características específicas entre la que podemos destacar:

- Uso de variables cualitativas. Normalmente, la población suele dividirse en clases en función de variables cualitativas como sexo, nivel cultural, orientación sexual, estado civil, etc. Incluso aquellas variables de naturaleza numérica como pueden ser edad, son convertidas a cualitativas ya que lo que se analiza es la pertenencia o no a un determinado rango de edad.

- Abundancia de intuiciones y carencia de conocimiento certero. Ante un determinado fenómeno, no resulta extraño que cualquiera de nosotros pueda formular una idea intuitiva de cual será el comportamiento de un determinado grupo social, aunque probablemente esta idea carecerá de la más mínima base científica o sociológica.

- La cantidad de información puede ser tan grande y referida a un número tal de elementos que nos resultará muy difícil analizarla y extraer algún tipo de relación entre variables de una manera directa.

En este tipo de problemas puede resultar muy beneficioso la realización de un análisis exploratorio de datos, de manera que más que obtener resultados muy concretos y precisos sobre el comportamiento de los distintos individuos que componen nuestra muestra de estudio, nos proporcione indicios sobre las relaciones entre ellos, que en manos de expertos en el problema en cuestión nos permita dirigir la investigación en una o otra dirección de búsqueda.

En este tipo de análisis nos parece de especial interés la herramienta computacional conocida como Mapas Auto-Organizativos de Kohonen (SOM) [14-15]. Esta técnica, encuadrada en el campo de la Inteligencia Computacional constituye uno de los más robustos clasificadores no supervisados y presenta una salida gráfica que facilita enormemente el análisis visual directo de las relaciones entre datos. En definitiva, está especialmente orientada al análisis de

datos difícilmente manejables por su complejidad, para los que no se conocen los criterios de clasificación, aunque se intuye que deben existir.

En este trabajo aplicamos la técnica SOM al análisis exploratorio de una base de datos formada por 999 patrones correspondientes a individuos anónimos que son portadores del virus VIH o han desarrollado SIDA en Cuba y cuya infección fue detectada entre 1990 y 1996.

Las variables cualitativas que describen a cada individuo aparecen en la tabla 2. Se han considerado 5 variables con un total de 15 modalidades.

Tabla 2. Variables cualitativas y sus modalidades en la base de datos utilizada

Variable	***Modalidades***
Edad	$mod_1 \equiv (13,20]$, $mod_2 \equiv (20,30]$, $mod_3 \equiv (30, \rightarrow)$
Sexo	$mod_4 \equiv$Hombre, $mod_5 \equiv$Mujer
Estado Civil	$mod_6 \equiv$Casado, $mod_7 \equiv$Soltero,
Nivel de Estudios	$mod_8 \equiv$Analfabeto, $mod_9 \equiv$Primaria, $mod_{10} \equiv$Secundaria, $mod_{11} \equiv$Bachiller, $mod_{12} \equiv$Técnico, $mod_{13} \equiv$Universitario
Orientación Sexual	$mod_{14} \equiv$HSH, $mod_{15} \equiv$HT

Esto significa que se dispone de 999 vectores binarios de 15 componentes. Así, p. ej., el vector $v = (0,0,1,1,0,1,0,0,0,0,1,0,1)$ representa a una persona mayor de 30 años (v_3=1), de sexo hombre (v_4=1), casado (v_6=1), con nivel académico de bachiller (v_{11}=1) y heterosexual (v_{15}=1).

La base de datos ha sido utilizada como entrada para el entrenamiento de una SOM de 10x10 neuronas mediante el siguiente algoritmo de entrenamiento estándar:

Paso 1: *Inicialización de variables*:
- Inicializar el radio de vecindad V y la razón de aprendizaje *a*
- Inicializar y normalizar los vectores de pesos de cada neurona

Paso 2: *Repetir hasta que se cumpla el criterio de convergencia*:
- Presentar un vector de entrada $Y^{(i)}$
- Obtener la neurona ganadora, es decir, aquella que cumpla: $\max_j (w_j \cdot Y^{(i)})$, siendo w_j el vector de pesos de la neurona j

- Modificar los vectores de pesos en la vecindad V de la neurona ganadora:

$$w_j(t+1) = w_j(t) + a(Y^{(i)} - w_j) \qquad (3)$$

- Normalizar los vectores de pesos
- Actualizar V y *a*

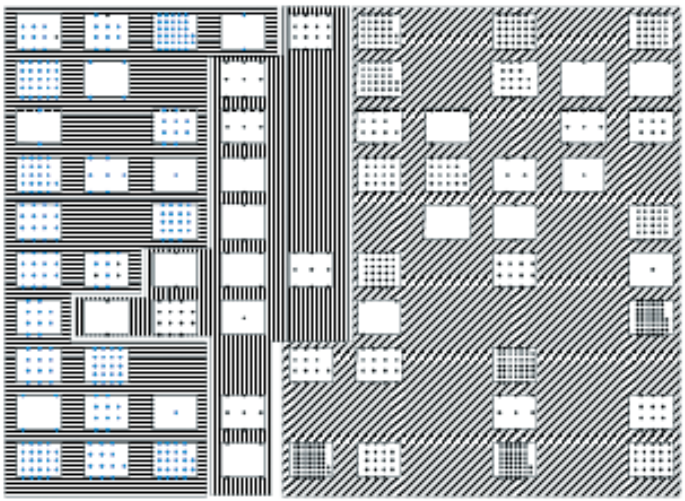
En nuestro caso, la razón de aprendizaje ha sido actualizada en cada iteración siguiendo la ecuación (14)

$$l_r(t) = l_{r0} / (1 + \frac{ct}{nn}) \qquad (14)$$

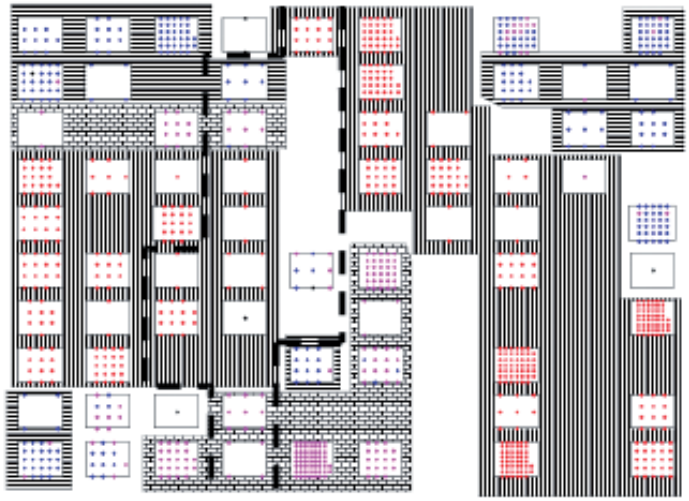
Donde l_{r0} es la razón de aprendizaje inicial (0.3 en nuestros experimentos), *c* es una constante (0.2), *t* es la iteración actual y *nn* es el número de neuronas de la red.

La figura 8 presenta la distribución de patrones en el mapa una vez concluido el proceso de clasificación no supervisado. Podemos destacar que un único mapa permite visualizar de manera simultánea la clasificación de los patrones por criterios de sexo, orientación sexual, y edad. Esta capacidad de visualización global del paradigma de Kohonen nos facilitará el análisis exploratorio de la incidencia de algún factor de la epidemia para cada clase poblacional obtenida.

En este trabajo, estamos interesados en analizar el tiempo que tarda el sistema sanitario cubano en detectar a una persona que está infectada por el virus VIH en función de la clase poblacional a la que esta persona pertenece. La disminución de este tiempo es un objetivo de alto interés ya que el conocimiento de su condición de seropositivo permitirá a esa persona adoptar pautas de conducta sexual más seguras, disminuyendo así su capacidad de infectar, así como recibir un tratamiento médico acorde con su situación que le permita mantener una calidad de vida más alta durante más tiempo. Sin embargo, es evidente que nuestra base de datos no puede proporcionarnos el dato del tiempo de demora en la detección, ya que no disponemos de la fecha en que una persona fue infectada. Sin embargo, si disponemos de otra variable temporal, el *tiempo de latencia* t_{lat}, definido como el tiempo transcurrido desde la fecha en que una persona es detectada como infectada (Ddet) hasta la fecha en que desarrolla los síntomas de la enfermedad (Dsida). Si partimos de la hipótesis de que una vez que la infección es detectada los cuidados paliativos son uniformes para toda la población, el tiempo de latencia puede dar una idea aproximada del retraso en la detección: cuanto mayor sea t_{lat}, más temprana será la detección y viceversa.



a)



b)

Figura 8. Organización final del mapa de Kohonen. a) Clasificación de acuerdo con las variables *Sexo* and Orientación Sexual (representa mujeres heterosexuales, representa hombres heterosexuales and representa hombres que practican sexo con hombres). b) Clasificación de acuerdo con la *Edad* (representa $13 \leq$ edad < 20, $20 \leq$ edad < 30, y $30 \leq$ edad).

Nosotros hemos *proyectado* la variable *tiempo de latencia* sobre el mapa de Kohonen, representando cada individuo mediante un símbolo particular en función de su tiempo de latencia. Así, en la figura 9, los puntos (•) representan individuos con un tiempo de latencia $t_{lat} \leq 3$, las aspas (x) representan individuos con $3 < t_{lat} \leq 7$, y los círculos (o) representan individuos con $7 < t_{lat}$. Mediante un análisis visual directo del mapa, podemos obtener las siguientes observaciones: en primer lugar, la proporción de aspas y círculos es mayor en la población de mujeres que en la de hombres; también, para la clase de mujeres aparece una mayor proporción de estos símbolos en el grupo de mayores de 30 años; con respecto a la clase de hombres, aspas y círculos son proporcionalmente más abundantes en la clase de Hombres que practican Sexo con Hombres.

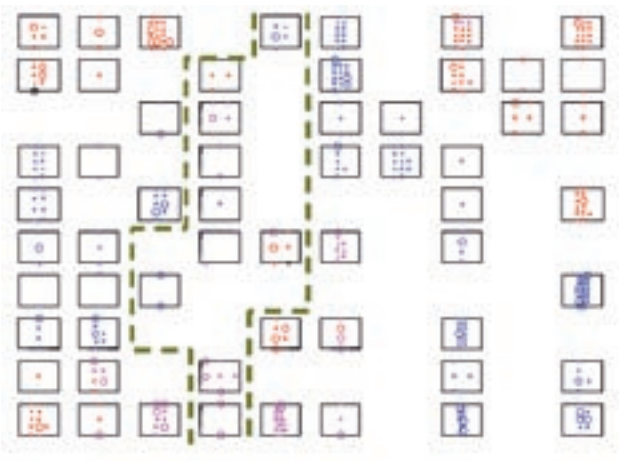


Figura 9. Representación de la clasificación de Kohonen de acuerdo al tiempo de latencia. Los puntos (•) representan individuos con $t_{lat} \leq 3$, cruces (x) representan $3 < t_{lat} \leq 7$, y círculos (o) representan $t_{lat} > 7$.

Estas percepciones cualitativas pueden ser confirmadas numéricamente por medio del diagrama de barras de la figura 10, el cual ha sido obtenido a partir de la base de datos con posterioridad. Efectivamente, la figura 10.a muestra una mayor proporción de individuos con $t_{lat} < 3$ entre los hombres heterosexuales que entre los HSH. Este hecho sugiere un mayor retraso en la detección para los primeros. La fugura 10.b ratifica el comportamiento especialmente de las mujeres mayores de 30 años: la proporción de individuos con $t_{lat} < 3$ es extraordinariamente bajo respecto a las demás.

Figura 10.- Distribución de la población de acuerdo al tiempo de latencia. De abajo a arriba, los diferentes bloques representan las proporciones de población con tiempos de latencia $t \leq 3$, $3 < t \leq 5$, $5 < t \leq 7$ y $t > 7$, respectivamente. a) Comparación entre hombres heterosexuales, hombres que practican sexo con hombres y mujeres. b) Distribución para mujeres $13 \leq$ edad< 20, $20 \leq$ edad< 30, y $30 \leq$ edad.

Comentarios a los resultados de análisis

Estos resultados parecen orientarnos en el siguiente razonamiento: por una parte, las políticas de detección parecen más eficientes en la clase de hombres que practican sexo con hombres que para la clase de hombres

heterosexuales; por otra parte, la detección parece especialmente efectiva para la clase de mujeres mayores de 30.

La presentación de estos datos a personas expertas en los aspectos sanitarios ha originado otras vías de interpretación. Así, el distinto comportamiento de las clases puede sugerir la posibilidad de que haya distintos virus VIH asociados a la infección en cada clase. Los resultados parecerían indicar que puede haber un número significativo de hombres heterosexuales que son infectados con cepas especialmente virulentas del virus, lo que explicaría un desarrollo más rápido de la enfermedad.

Los resultados anteriores indican por un lado, la extraordinaria eficacia de los Mapas Auto-Organizativos de Kohonen como herramientas para el análisis exploratorio de datos mediante un proceso de visualización directa. En el contexto que nos ocupa puede ser de una gran utilidad en la orientación de los expertos sanitarios en el conocimiento de muchos factores de interés para el desarrollo de la epidemia y su control.

Finalmente, respecto a la variable de nivel de estudios, es necesario decir que no se ha producido una clasificación clara por este concepto en la red, lo que viene a indicar que esta variable no es significativa en la epidemia, es decir, que el nivel de estudios no constituye ninguna garantía frente a esta epidemia.

Capítulo V.
Conclusiones

En esta memoria se resume el trabajo llevado a cabo por nuestro grupo de investigación en relación al estudio de la epidemia VIH-SIDA en Cuba. Partimos de un modelo matemático de dicha epidemia basado en Ecuaciones Diferenciales Ordinarias, lo que supone un enfoque poco común en epidemiología, aunque altamente productivo.

El trabajo presenta, a nuestro entender, un doble interés:

En primer lugar, desde el punto socio-sanitario, nos ha permitido realizar un acercamiento al conocimiento de epidemia VIH-SIDA en Cuba desde una triple perspectiva: análisis de la dinámica y evolución de la epidemia, estimación de los parámetros que intervienen en su modelo, y análisis exploratorio de su incidencia para distintos grupos de población.

A partir del análisis dinámico hemos aportado información matemáticamente fundamentada sobre la situación actual y tendencia de evolución de la epidemia, y se ha valorado cuantitativamente el esfuerzo que seria necesario para detener su crecimiento. Así, a partir de nuestra estimación de parámetros hemos calculado el Número Básico de Reproducción de la epidemia R_0, resultando de 1.34 con un intervalo de confianza (1.31, 1.36). Este valor demuestra que la epidemia, aunque en fase de crecimiento, está muy controlada en Cuba, ya que el valor que los expertos dan a R_0 para una enfermedad de transmisión sexual como el VIH-SIDA está en el intervalo (2, 5). Así mismo se han calculado los valores de k_1 y k_2, parámetros relacionados con la tasa de detección de nuevos infectados por medios azarosos y por búsqueda activa de contactos respectivamente, que serían necesarios para que la epidemia evolucionase a su erradicación. Estos valores supondrían un aumento del 25% en la tasa de detección azarosa(k_1) y del 5% en k_2. Ambas condiciones estan lejos de poder ser alcanzadas cen las condiciones económicas de Cuba, por tanto, la conclusión principal en este punto es que la epidemia VIH-SIDA no podrá erradicarse ni siquiera llegar a un estado de endemia estacionario, pero que si se mantiene el esfuerzo actual por parte del sistema sanitario cubano, su crecimiento será muy moderado.

Respecto a la identificación del sistema, se ha conseguido la estimación de los parámetros más significativos de la epidemia, como los ya mencionado k_1 y k_2, relacionados con las tasas de detección de nuevos infectados por medios azarosos y por búsqueda activa de contactos respectivamente, β, la tasa de infectados desconocidos que acaban desarrollando la enfermedad sin haber sido detectados como tales (inversa del tiempo de incubación de la enfermedad), λ y λ', tasas de infección producida por seropositivos no detectados y detectados respectivamente. La estimación de estos parámetros tiene una gran importancia

social y económica, ya que proporcionan información sobre la eficiencia del programa de lucha contra la epidemia desarrollado por el sistema sanitario de un país; pero también tiene un valor añadido desde el punto de vista matemático y computacional, debido a la dificultad encontrada en trabajos anteriores para llevarla a cabo a través de métodos clásicos de estimación paramétrica. En este trabajo se han utilizado dos métodos englobados en lo que se conoce como técnicas bio-inspiradas o de inteligencia computacional: Algoritmos Genéticos y Redes Neuronales Artificiales. Los resultados obtenidos en ambos casos concuerdan con las estimaciones manejadas por los expertos.

Finalmente, el análisis exploratorio de los datos de la epidemia nos ha permitido llevar a cabo una clasificación directamente visualizable de las personas infectadas en función de un conjunto de variables cualitativas como sexo, tendencia sexual, estado civil, rango de edad y nivel de estudios. A continuación se ha analizado la incidencia del tiempo de latencia de la enfermedad (tiempo transcurrido desde que una persona es detectada como portadora hasta que desarrolla SIDA) sobre cada uno de los grupos poblacionales obtenidos. Como resultados más relevantes, se ha detectado un comportamiento especialmente interesante de las mujeres mayores de 30 años respecto a esta variable, ya que presentan un tiempo de latencia mucho mayor que los demás grupo. Esto sugiere la detección de la enfermedad en estas mujeres es más temprana que en el resto. Por otra parte, se observa un menor tiempo de latencia en los hombres heterosexuales (HT) que en aquellos que practican sexo con otros hombres (HSH). Este fenómeno puede tener distintas líneas de explicación: por una parte, pudiera pensarse que el sistema sanitario cubano detecta más fácilmente a los HSH, pero por otra parte, pudiera pensarse que un la infección de un número significativo de HT haya sido producida por una cepa o combinación de cepas especialmente virulentas. Este análisis exploratorio ha sido llevado a cabo mediante otra técnica de inteligencia computacional: los Mapas Auto-Organizativos de Kohonen.

Por otra parte, desde un punto de vista científico-computacional, como ya ha sido esbozado, el uso de las técnicas computacionales mencionadas, relacionadas con los métodos Bio-inspirados (Algoritmos Genéticos) y de Inteligencia Computacional (Redes Neuronales Recurrentes de Hopfield y Mapas Auto-Organizativos de Kohonen) nos ha permitido presentar soluciones novedosas y eficaces al problema de la estimación paramétrica, para el que no contábamos con métodos analíticos o numéricos clásicos totalmente satisfactorios. Así mismo, este trabajo ha permitido describir de manera meticulosa y ordenada los fundamentos y la metodología de aplicación de estas técnicas, tanto en problemas generales de optimización como en problemas de clasificación y visualización de datos. Es necesario destacar que la aplicación de estas técnicas

computacionales puede considerarse aún en fase de emergencia en un gran número de los centros de investigación de Hispanoamérica.

Como productos “marginales” de esta investigación, se han obtenido dos paquetes software destinados al diseño asistido de algoritmos genéticos para el problema de la estimación paramétrica de un sistema EDO y para la clasificación no supervisada y el análisis exploratorio de datos mediante Mapas Auto-Organizativos de Kohonen.

Agradecimientos

Este trabajo contó con el apoyo de la Agencia Española de Cooperación Internacional con sus proyectos A/2051/04, A/2840/05 y A/6294/06.

Los autores quieren agradecer al Departamento de Tecnología Electrónica de la Universidad de Málaga por la acogida y el apoyo material, al darnos un espacio en su seno donde poder desarrollar estos trabajos en colaboración con los miembros del equipo de Investigación de Ingeniería de Sistemas Integrados (ISIS), (TIC-125), de dicha universidad, con todas las condiciones óptimas de trabajo que pusieron a nuestra disposición.

En especial agradecemos al Dr. Gonzalo Joya Caparrós quien como coordinador de los mencionados proyectos AECI nos alentó a realizar estas investigaciones y nos dio todo su apoyo en todo momento. También a los Doctores Miguel Atencia y Francisco García Lagos por introducirnos en las técnicas de las redes neuronales recurrentes tipo Hopfield y los mapas de Kohonen. También a la Ingeniera Maria Esther García Garaluz por su ayuda en los programas de cómputo.

Referencias

[1] Ying-Heng Hsieh, de Arazoza, Hector y otros, A class of methods for HIV contact tracing in Cuba: implications for intervention and treatment. En: HYPERLINK "http://www.ams.org/mathscinet/search/publications.html?pg1=IID&s1=170555" Tan, Wai-Yuan; HYPERLINK "http://www.ams.org/mathscinet/search/publications.html?pg1=IID&s1=629585" Wu, Hulin, editores, Deterministic and stochastic models of AIDS epidemics and HIV infections with intervention. *World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ,* 2005. viii+601 pp. ISBN: 981-256-139-0.

[2] Joint United Nations Programme on HIV/AIDS and World Health Organization, *AIDS Epidemic Update: December 2000*. Ginebra: UNAIDS/WHO, Diciembre 2000. *http://www.unaids.org/wac/2000/wad00/files/WAD_epidemic_report.PDF*

[3] WHO, U.-. *AIDS epidemic update: December 2005.*

[4] Torres Peña, R. y otros, *Prevention and Control of HIV/AIDS Program. The Cuban Experience.* Ministerio de Salud Pública de Cuba. 2002. *HYPERLINK "http://www.hiv-lac-epinet.org" http://www.hiv-lac-epinet.org*

[5] Lounes, R. y de Arazoza Hector, *A two type model for the Cuban National Program on HIV/AIDS,* IMA J. Math. App. Med. Biol. 16 (1999), 143-154

[6] de Arazoza, H. y Lounes, R. *A non linear model for a sexually transmited disease with contact tracing.* Mathematical Medicine and Biology: A Journal of the IMA Vol. 19, No. 3, 2002.

[7] Luenberger, D.E., Linear and Nonlinear Programing, Second Edition. Addison-Wesley Publishing Co, Inc, Massachusets, E.U.A., 1984

[8] Dennis, J. E., Schnabel, R.B., Numerical methods for unconstrained optimization and non linear equations. Prentice-Hall Series in Computational Mathematics. New Jersey, 1983

[9] Michalewicz, Z., Genetic SAlgorithms + Data Structures = Evolution Programs, Springer, 3ª Edición ,1996

[10] Hopfield, J., Neural Networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA 79 (1982) 2554–2558.

[11] Atencia, M., Joya, G., Sandoval, F., Optimización con Sistemas Neuronales de Hopfield. Aplicación a la Identificación Paramétrica de Sistemas Dinámicos. En *Optimización Inteligente. Técnicas de Inteligencia Computacional para Optimización*, Gonzalo Joya, Miguel Atencia, Alberto Ochoa y Sira Allende (Coordinadores), 2004, Servicio de Publicaciones e Intercambio Científico de la Uiversidad de Málaga, pp. 3-42

[12] Atencia, M., Joya, G., Sandoval, F., Dynamical analysis of continuous higher order Hopfield networks for combinatorial optimization, *Neural Computation*, **vol. 17,** pp. 1802-1819, 2005.

[13] Abe, S., Theories on the Hopfield Neural Networks. In: Proc. IEE International Joint Conference on Neural Networks. Volume I. (1989) 557–564.

[14] Haykin, S., *Neural Networks. A Comprehensive Foundation*. Macmillan College Publishing Company Inc., New York, 1994.

[15] Kohonen, T., The Self-Organizing Map, *Proceedings of the IEEE* , **vol. 9, no.** 78, pp. 1464-1480, (1990)

www.unia.es

Universidad Internacional de Andalucía

Premio de estudios Iberoamericanos Grupo La Rábida 2

Área Científico-Técnica