



UNIVERSIDAD  
DE MÁLAGA

Programa Oficial de Posgrado en Telecomunicación  
Escuela Técnica Superior de Ingeniería de Telecomunicación

*TESIS DOCTORAL*

Modelado matemático de sistemas  
dinámicos en epidemiología

AUTORA: MARÍA ESTHER GARCÍA GARALUZ

DIRECTORES: MIGUEL ALEJANDRO ATENCIA RUIZ  
GONZALO JOYA CAPARRÓS

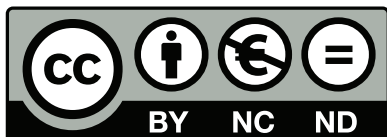
MÁLAGA, 2014



**Publicaciones y  
Divulgación Científica**

AUTOR: María Esther García Garaluz

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está sujeta a una licencia Creative Commons:

Reconocimiento - No comercial - SinObraDerivada (cc-by-nc-nd):

[Http://creativecommons.org/licenses/by-nc-nd/3.0/es](http://creativecommons.org/licenses/by-nc-nd/3.0/es)

Cualquier parte de esta obra se puede reproducir sin autorización pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): [riuma.uma.es](http://riuma.uma.es)

D. Miguel A. Atencia Ruíz, Profesor Titular del Departamento de Matemática Aplicada de la Universidad de Málaga y D. Gonzalo Joya Caparrós, Profesor Titular del Departamento de Tecnología Electrónica de la Universidad de Málaga, como directores de la Tesis Doctoral de María Esther García Garaluz,

CERTIFICAN:

Que María Esther García Garaluz, Ingeniera de Telecomunicación y Máster en Tecnologías de Telecomunicación, ha realizado en el Departamento de Matemática Aplicada y en el Departamento de Tecnología Electrónica de la Universidad de Málaga, bajo su dirección, el trabajo de investigación correspondiente a su Tesis Doctoral titulado:

## **Modelado matemático de sistemas dinámicos en epidemiología**

Revisado el presente trabajo, estiman que puede ser presentado al Tribunal que ha de juzgarlo. Y para que conste, autorizan la defensa de esta Tesis Doctoral en la Universidad de Málaga.

Málaga, a 22 de Abril de 2014



*A mi familia*



# Agradecimientos

Me gustaría expresar mi agradecimiento a Miguel Atencia y Gonzalo Joya por la ayuda y dedicación en la realización de esta tesis, y por todas las cosas que me han dado la oportunidad de realizar en este tiempo. Sin ellos, sin duda, no habría recorrido este camino.

A Madalina Olteanu, junto con el resto del equipo del grupo SAMM, y al equipo de la Facultad de Matemática y Computación de la Universidad de la Habana, por su amabilidad y disponibilidad durante mis estancias en Francia y Cuba.

A mi familia, por ofrecerme siempre su cariño y apoyo. A Salva, por su implicación y por hacer que las cosas sean siempre más fáciles.

A mis compañeros de trabajo, por su generosidad cubriendo mis ausencias, especialmente durante mi estancia.

Y a todos aquellos, que de una forma u otra, me ayudan a realizar mis proyectos.





# Resumen

El trabajo presentado en esta tesis se centra en la identificación de sistemas dinámicos mediante técnicas de inteligencia computacional. Concretamente, se realiza el modelado y estimación de parámetros de las epidemias de VIH/SIDA y dengue para el caso de Cuba. Para la estimación de los parámetros de los modelos presentados se utiliza un método de estimación basado en las redes neuronales de Hopfield, inspirado por su capacidad para resolver problemas de optimización.

En el caso de la epidemia de VIH/SIDA, el trabajo realizado parte de un modelo de ecuaciones diferenciales ordinarias que describe la dinámica de poblaciones de la infección en Cuba. El modelo contiene una serie de parámetros que ponderan las transiciones entre poblaciones, entre los que destacan los parámetros que describen la detección de individuos seropositivos por el sistema sanitario del país. Los distintos métodos de detección se agrupan de manera que se incluyen de forma explícita solo dos grupos: la detección mediante el programa de notificación de parejas, un sistema de búsqueda activa de seropositivos en el que el sistema sanitario ha empleado un especial esfuerzo con medios materiales y humanos, y la detección mediante el resto de métodos. Un objetivo de este trabajo es permitir el análisis de la eficacia del programa de búsqueda activa de seropositivos. Ambos parámetros de detección son muy difíciles de estimar mediante métodos clásicos, por lo que en este trabajo se propone su estimación mediante el método basado en redes neuronales de Hopfield. Este método había sido probado con buenos resultados en la estimación de parámetros de modelos mecánicos, pero no había sido utilizado con anterioridad en modelos discretos y sin entradas, como es el caso de los sistemas

epidemiológicos. También se ha definido un umbral de detección de la epidemia, que determina el estado de crecimiento o decrecimiento de la epidemia, de manera similar al número básico de reproducción definido en el ámbito de la epidemiología. Finalmente, se han calculado las poblaciones predichas de la epidemia a partir de los resultados de la estimación y de una base de datos reales de la epidemia en Cuba. Los resultados obtenidos, tanto para la estimación de los parámetros como para el umbral de detección, son coherentes con los obtenidos en otros trabajos y la comparación de las poblaciones predichas con las reales es satisfactoria y válida la bondad tanto del modelo como de la propia estimación.

La mayoría de los modelos del VIH/SIDA propuestos, incluido el descrito en el párrafo anterior, cuentan entre sus variables con la población de individuos infectados de VIH que no han sido detectados aún por las autoridades sanitarias. Esto resulta una seria limitación a la hora de validar los resultados comparando con datos reales. Por ello, en esta tesis se propone también un nuevo modelo para el VIH/SIDA en Cuba en el que no aparece la población infectada no detectada, sino que se define una ratio de infección oculta como el cociente de los individuos infectados de VIH desconocidos entre los conocidos. A partir de esta definición y del modelo del VIH/SIDA previo, se define un nuevo modelo que dependerá únicamente de la ratio de infección oculta, permitiendo obtener información, de manera indirecta, sobre la población de seropositivos no detectados. Para el modelo propuesto se han realizado los análisis de estabilidad y de sensibilidad respecto a los parámetros. En este caso, también se ha estudiado el ajuste del modelo a la epidemia utilizando una base de datos reales de la epidemia en Cuba. Los resultados obtenidos muestran un buen ajuste a los datos reales y se corresponden con la historia socio-económica de la epidemia y del país, además de dar una idea de las posibles mejoras y extensiones del modelo.

Respecto a la epidemia de dengue en Cuba, en primer lugar se define un modelo paramétrico que describe la dinámica de poblaciones de la epidemia mediante un sistema de ecuaciones diferenciales con

---

retardo. La utilización de ecuaciones con retardo evita la necesidad de incluir explícitamente en el modelo las poblaciones de mosquitos vectores del virus. Esta es una de las principales aportaciones de este modelo, ya que los estudios con poblaciones de mosquitos no pueden validarse directamente con datos reales. Por su parte, desde el punto de vista matemático, la aplicación del estimador basado en redes neuronales de Hopfield a ecuaciones con retardo representa una interesante novedad. Para este modelo se ha definido también un umbral de detección que determina el estado de evolución de la epidemia. Los experimentos numéricos muestran un ajuste satisfactorio entre la población predicha de individuos infectados y los datos reales. Asimismo, los resultados de la estimación y del umbral de detección sugieren que la extensión de la epidemia sería más prolongada y severa sin políticas de fumigación. En conclusión, el trabajo presentado apoya la idea del uso del estimador en ecuaciones con retardos y permite evaluar la eficacia de las medidas de control de la epidemia.



# Abstract

The work that is introduced in this thesis focuses on identification of dynamical systems using computational intelligence techniques. In particular, we perform modelling and parameter estimation of the HIV/AIDS and dengue epidemics in the case of Cuba. In order to estimate the parameters of the presented models, we use an estimation method based on Hopfield neural networks, inspired by their ability to solve optimization problems.

In the case of the HIV/AIDS epidemic, the work stems from a model that describes the population dynamics of the epidemic in Cuba. The model is defined by ordinary differential equations, and it contains a number of parameters that weight the transitions between populations. Among these parameters, those that describe the detection of seropositive individuals by the health system of the country stand out. The different detection methods are grouped so as to explicitly include only two sets: detection by the partner notification program, an active search system for HIV infected individuals on which the health system has devoted a substantial effort with human and material resources, and detection by other methods. One of the objectives of this work is to provide means for an analysis of the effectiveness of the active search for seropositives. Both detection parameters are very difficult to estimate by conventional methods, so we propose using the method based on Hopfield neural networks as estimator. This method has been tested with satisfactory results in the estimation of parameters of mechanical models, but has not been applied on discrete models without inputs, as is the case of epidemiological systems. We have also defined a detection threshold of the epidemic, which determines the state of

growth or decline of the epidemic, similarly to the basic reproduction number defined in epidemiology. Finally, we have computed the predicted populations of the epidemic from the estimation results and a base of real data of the epidemic in Cuba. The results, both for the parameter estimation and the detection threshold, are consistent with those from other contributions, and the fit of the predicted populations to the real ones is satisfactory, thus providing a positive assessment of both the goodness of the model and the estimation itself.

Most HIV/AIDS models, including that described in the previous paragraph, include as one of their variables the population of HIV-infected individuals who have not yet been detected by the health authorities. This is a severe drawback as long as results should be compared to real data. Therefore, in this thesis we also propose a new model for HIV/AIDS in Cuba, which does not explicitly include the undetected infected population but, instead, a ratio of hidden infection is defined as the quotient of unknown infected individuals by those known. Starting from this definition and the previous HIV/AIDS model, a new model that only depends on the ratio of hidden infection is defined, which indirectly provides information on the population of undetected seropositive individuals. For the proposed model, the stability analysis and the study of sensitivity to parameters are performed. In this case, we have also studied the fitting of the model to real data using a database of the epidemic in Cuba. The obtained results show a satisfactory agreement to real data, and they are consistent with the socio-economic history of the epidemic and the country, besides providing an indication of the possible improvements and extensions of the model.

Regarding the epidemic of dengue in Cuba, a parametric model describing the population dynamics of the epidemic through a system of delay differential equations is defined. Using delay equations overcomes the need for explicitly including in the model the populations of mosquitoes that are responsible for the contagion. This is a major contribution of this model, since studies with mosquito populations cannot be validated using real data. On another note, from

---

the mathematical point of view, the application of the estimator based on Hopfield neural networks to delay equations is an interesting contribution. For this model we have also defined a detection threshold that determines the evolution status of the epidemic. Numerical experiments show a satisfactory adjustment between the predicted population of infected individuals and real data. Likewise, the results of the estimation and the computation of the detection threshold suggest that the epidemic spread would be longer and more severe if it was not for fumigation policies. As a conclusion, the presented work supports using the estimator on delay differential equations and provides an assessment on the efficacy of epidemiological control measures.





# Índice general

<b>1. Introducción</b>	<b>17</b>
<b>2. Estimación de parámetros en la epidemia de SIDA en Cuba</b>	<b>25</b>
2.1. Introducción . . . . .	25
2.2. Estimación de parámetros mediante redes neuronales de Hopfield . . . . .	26
2.3. Modelo de la epidemia de VIH/SIDA en Cuba . . . . .	29
2.4. Resultados experimentales . . . . .	31
2.4.1. Estimación de parámetros de detección . . . . .	32
2.4.2. Cálculo de tasa de reproducción de la epidemia . . . . .	34
2.4.3. Predicción de la población infectada . . . . .	36
2.5. Conclusiones . . . . .	37
<b>3. Ratio de la infección oculta del VIH en Cuba</b>	<b>39</b>
3.1. Introducción . . . . .	39
3.2. Modelo de la ratio de infección oculta de VIH en Cuba . . . . .	40
3.2.1. Análisis de estabilidad . . . . .	42
3.2.2. Análisis de sensibilidad . . . . .	43
3.3. Pruebas con datos reales . . . . .	47
3.4. Conclusiones . . . . .	50
<b>4. Identificación de un modelo de dengue en Cuba</b>	<b>53</b>
4.1. Introducción . . . . .	53
4.2. Modelo de la fiebre de dengue en Cuba . . . . .	54
4.2.1. La fiebre de dengue . . . . .	54

4.2.2. Definición del modelo de dengue en Cuba . . . . .	54
4.2.3. Número básico de reproducción . . . . .	57
4.3. Estimador de Hopfield en sistemas con retardo . . . . .	58
4.4. Pruebas con datos reales . . . . .	59
4.5. Conclusiones . . . . .	62
<b>5. Conclusiones</b>	<b>65</b>
<b>6. Conclusions</b>	<b>75</b>
<b>Bibliografía</b>	<b>83</b>
<b>A. Publicaciones que compendian la tesis</b>	<b>89</b>
A.1. Estimation of parameters based on artificial neural networks and threshold of HIV/AIDS epidemic system in Cuba . . . . .	89
A.2. The ratio of hidden HIV infection in Cuba . . . . .	105
A.3. Hopfield networks for identification of delay differential equations with an application to dengue fever epidemics in Cuba . . . . .	127

# Capítulo 1

## Introducción

El problema general del que trata esta tesis es la construcción, validación, análisis e identificación de sistemas dinámicos que modelan epidemias. Una característica crítica de los modelos epidemiológicos, en comparación, por ejemplo, con los sistemas mecánicos, es que no se dispone de un modelo que emane de forma precisa de leyes físicas. Por el contrario, los modelos epidemiológicos, cuando existen, son el resultado intuitivo del criterio de expertos, conteniendo siempre un margen de incertidumbre. En este caso, la identificación de sistemas puede definirse como la determinación de las propiedades internas del sistema a partir de sus salidas externas observadas [32], y se integra íntimamente en el proceso de modelado: la obtención de resultados satisfactorios de identificación sugiere la validez del modelado, mientras que una identificación defectuosa puede deberse a un modelado inexacto. De esta forma, en el proceso de identificación se pueden distinguir varias tareas [4]:

- Establecer el dominio del modelo, ya sea en el tiempo o en la frecuencia. Si el modelo se formula en el dominio del tiempo, de forma general podrá definirse mediante ecuaciones diferenciales, que describirán la evolución del estado del sistema. Este trabajo se centra en modelos descritos por Ecuaciones Diferenciales Ordinarias, pudiendo incluir retardos. Los estados de los modelos epidemiológicos suelen representar *poblaciones* (por ejemplo, de

individuos sanos, infectados, etc.).

- Una vez establecida la estructura básica del modelo, en este suelen aparecer parámetros, es decir, magnitudes cuyo valor mide la ocurrencia de un determinado proceso. Centrándonos en el caso epidemiológico, los parámetros corresponden habitualmente a tasas de transferencia entre poblaciones por unidad de tiempo (por ejemplo, infección, mortalidad, etc.).
- Finalmente, el modelo quedará totalmente concretado al establecer el valor numérico de los parámetros que contiene. El proceso de cálculo de los parámetros constituye la tarea denominada *estimación* [40] o identificación paramétrica. Desde un punto de vista matemático y computacional, la estimación de parámetros es la técnica fundamental del presente trabajo.

En consecuencia, cabe realizar algunas consideraciones sobre el estado de la cuestión del modelado de epidemias, por una parte, y de la estimación de parámetros, por otra.

Las enfermedades infecciosas son uno de los problemas de mayor transcendencia para las autoridades sanitarias, que periódicamente plantean nuevas estrategias para mejorar la prevención, detección temprana y tratamiento de los afectados, y para establecer medidas de control de su evolución. El estudio y modelado matemático de enfermedades infecciosas es un problema que ha sido abordado frecuentemente a lo largo de la historia para ayudar a plantear esas estrategias [15, 13, 19, 33]. En este contexto, existen varios trabajos interesantes en la literatura que analizan las epidemias estudiadas aquí: la epidemia de VIH/SIDA y la de dengue. A menudo, un objetivo fundamental de los trabajos de modelado epidemiológico es el cómputo o la estimación del *número básico de reproducción* [18], que es un valor numérico que caracteriza la evolución futura de la epidemia: si este valor es mayor que la unidad, se predice que la enfermedad se expandirá de forma indefinida, mientras que el caso contrario indica que la infección es autolimitada y la epidemia terminará desapareciendo.

Respecto a la epidemia de VIH/SIDA, su aparición supuso un gran impacto social debido a su gravedad y despertó interés por su análisis matemático desde el principio. En la literatura se han propuesto distintos modelos de la epidemia en función de los factores que se pretendían tener en cuenta o las técnicas utilizadas ([2, 28, 38, 27, 35, 34], entre otros). El estudio del VIH/SIDA en el caso de Cuba es especialmente interesante ya que la enfermedad presenta en este país una incidencia muy baja, especialmente si se compara con el resto de países del Caribe. En este trabajo se utilizará como punto de partida un modelo de la epidemia descrito en [27], que agrupa a las personas afectadas por la enfermedad en distintas poblaciones en función de si su infección ha sido detectada o no (es decir, registrada por el sistema sanitario) y, si lo ha sido, del método de detección utilizado. El objetivo de este modelo es estudiar la eficacia de los métodos de detección empleados y, en relación con ellos, la posible evolución de la enfermedad. Una característica crítica de la infección por VIH/SIDA es su largo y extremadamente variable periodo de incubación, que da lugar a un número significativo de individuos contagiosos que, al ser asintomáticos, desconocen que pueden propagar la infección. Desde un punto de vista epidemiológico, esta característica supone un obstáculo para validar los modelos propuestos, ya que el tamaño de la población de infectados *ocultos* es desconocida. Este será un aspecto muy presente a lo largo de esta tesis y, en concreto, una de las propuestas es la construcción de un modelo que no contiene explícitamente la población de infectados no detectados.

Por su parte, la incidencia de la fiebre de dengue ha crecido considerablemente en las últimas décadas y es una de las principales causas de mortalidad en algunos países de Asia y Latinoamérica [42]. Existen numerosos estudios en la literatura que tienen como objetivo modelar la enfermedad de dengue ([21, 29, 16], sin ánimo de ser exhaustivos). El dengue es una infección vírica que se transmite por la picadura de una especie particular de mosquitos, que actúan como vectores de la enfermedad. La mayoría de los modelos propuestos incluyen entre sus variables una o más poblaciones de mosquitos responsables de

transmitir la enfermedad. Dado que no es realista plantear medidas mínimamente precisas de las poblaciones de mosquitos (cuanto menos determinar la proporción de infectados), no es posible validar los modelos comparándolos con datos reales del desarrollo de la enfermedad. La propuesta de este trabajo es la construcción de un modelo que no hace referencia expresa a la población de insectos.

En la computación clásica existen diferentes métodos para estimar los parámetros de un modelo, que pueden ser clasificados en dos grandes grupos: los métodos de gradiente [37] y los métodos de mínimos cuadrados [36]. Mientras que los métodos de este último grupo no se ajustan bien a la estimación de parámetros variables en el tiempo, los primeros tienen el inconveniente de necesitar que el término de la ganancia se ajuste de manera empírica en cada caso concreto. Para solventar este problema se han propuesto múltiples alternativas (véase, por ejemplo, [3] para una revisión reciente), pero dista de poderse considerar un problema cerrado y sigue siendo necesaria la propuesta de contribuciones novedosas que presenten mejoras en sistemas específicos. A este respecto, cabe señalar que la mayoría de los métodos de estimación de parámetros se han validado rigurosamente al aplicarlos a sistemas mecánicos, por ejemplo, en el ámbito de la ingeniería de control. En cambio los sistemas epidemiológicos tienen una serie de características que los diferencian de los modelos físicos. Por un lado, sus modelos no se basan en leyes físicas sino en el conocimiento de expertos, por lo que el propio modelo contiene un margen de *incertidumbre* que no debe ser atribuido en exclusiva al método de estimación. Otra característica importante es que los sistemas no presentan *excitación externa*, es decir, no tienen entradas sobre las que se pueda actuar. Esta es una diferencia significativa con los sistemas de control, donde la definición de una entrada adecuada puede facilitar el proceso de identificación. Finalmente, los actuales sistemas mecánicos disponen de sensores que proporcionan medidas con periodos en el rango de los microsegundos, por lo que las señales involucradas pueden considerarse continuas en la práctica, y ser manejadas con las técnicas del cálculo diferencial. En cambio, en el caso de trabajar con datos epidemioló-

gicos reales, habrá que considerar que los modelos correspondientes son *discretos*, ya que los datos de los que se dispondrá no tendrán una periodicidad suficiente para poder considerarlos continuos. En todos estos sentidos, el estimador que se utiliza en este trabajo es una aportación en la dirección de proponer métodos que estimen parámetros variables en este tipo de modelos. Este estimador está basado en una formulación de las redes neuronales de Hopfield para la resolución de problemas de optimización, denominada formulación de Abe [1, 30, 10]. Este método, que ha sido utilizado previamente para la estimación de parámetros de sistemas mecánicos [9, 11], será utilizado para la estimación de parámetros definiendo el problema de estimación como un problema de optimización del error de predicción.

La estructura de la presente tesis se describe a continuación. En el capítulo 2 se muestra un modelo de la epidemia de VIH/SIDA en Cuba [27], basado en Ecuaciones Diferenciales Ordinarias (EDO). Asimismo, el método de estimación mediante redes neuronales de Hopfield será aplicado aquí para la estimación de los parámetros de este sistema, prestando especial atención a los parámetros que modelan los métodos de detección de nuevos infectados. La idea es obtener, de forma indirecta, información sobre la población *oculta* de individuos infectados. En esta misma línea, en el capítulo 3 de esta tesis se propone un nuevo modelo para la epidemia de VIH/SIDA en Cuba que no depende de las propias poblaciones sino de la razón entre los infectados desconocidos y los conocidos, lo que hemos dado en llamar *ratio de infección oculta*. El modelo obtenido sugiere que esta razón tiende asintóticamente a un valor constante, lo que muestra un acuerdo significativo con los datos existentes. Finalmente, en el capítulo 4 se presenta un modelo de epidemia de dengue en Cuba definido mediante Ecuaciones Diferenciales con Retardo (EDR). En este caso, la población de mosquitos no consta explícitamente en el sistema, siendo sustituido por un término con retardo que modela implícitamente todo el proceso por el que un individuo infecta a otro, incluyendo la transferencia del virus mediante el vector.

En resumen, la tesis aquí presentada conforma un trabajo multidisciplinar con los siguientes objetivos desde el punto de vista epidemiológico y computacional:

- Modelado de los sistemas dinámicos de VIH/SIDA y dengue.
  - Modelado de la ratio de infección oculta mediante ecuaciones diferenciales ordinarias.
  - Análisis y validación del modelo de la ratio de infección oculta.
  - Modelado de la epidemia de dengue mediante ecuaciones diferenciales con retardo.
  - Definición de los números básicos de reproducción del VIH/SIDA y del dengue en Cuba.
- Evaluación de la bondad de la estimación de parámetros mediante el método basado en redes neuronales de Hopfield.
  - Estimación de parámetros en modelos descritos por ecuaciones diferenciales ordinarias.
  - Estimación de parámetros en modelos descritos por ecuaciones diferenciales con retardo.

El trabajo presentado en esta tesis se ha desarrollado gracias a la colaboración del grupo ISIS del Departamento de Tecnología Electrónica de la Universidad de Málaga y de la Facultad de Matemática y Computación de la Universidad de la Habana en el marco de sucesivos proyectos de investigación de la Agencia Española de Cooperación Internacional para el Desarrollo (AECID). Con respecto a estos últimos, en concreto, la autora ha participado en los proyectos denominados D/009842/07, D/017218/08, D/023835/09 y D/030223/10. Como fruto de esta colaboración se ha definido el modelo de partida para la epidemia de VIH/SIDA en Cuba y el modelo de dengue. Los datos utilizados para la realización de las pruebas experimentales descritas en este trabajo se corresponden con dos bases de datos reales anónimas de las epidemias de VIH/SIDA y dengue en Cuba. La fuente



de dichas bases de datos es el Instituto de Medicina Tropical *Pedro Kourí* de Cuba. Como resultado directo de esta tesis se ha obtenido la publicación de diversos trabajos ([7, 22, 5, 6], entre otros). De estas publicaciones, tres corresponden a revistas con factor de impacto significativo según el listado *Journal Citation Reports*, y se han incluido en un apéndice del presente trabajo.



## Capítulo 2

# Estimación de parámetros en la epidemia de SIDA en Cuba

### 2.1. Introducción

En este capítulo se expone el trabajo descrito en el artículo “Estimation of parameters based on artificial neural networks and threshold of HIV/AIDS epidemic system in Cuba” [5]. El objetivo principal es la estimación de los parámetros presentes en un modelo de la epidemia VIH/SIDA en Cuba previamente descrito en la literatura [27]. El modelo está definido mediante Ecuaciones Diferenciales Ordinarias (EDO) y el método utilizado para la estimación de sus parámetros está basado en las redes neuronales de Hopfield, aprovechando la capacidad de este tipo de redes para resolver problemas de optimización. Además de la identificación del sistema de VIH/SIDA, los resultados de este trabajo también permiten realizar una validación del modelo propuesto. Finalmente, se calculará un umbral de detección, cuyo significado epidemiológico es análogo al número básico de reproducción, es decir, proporciona información de la evolución de la enfermedad. Para todas las pruebas realizadas, esto es, la estimación de parámetros, la validación del modelo y el cálculo del número básico de reproducción, se

utilizará una base de datos reales de la epidemia de 20 años (entre 1986 y 2006) en la que se registra la población de personas que ha desarrollado SIDA y las distintas poblaciones de personas infectadas de VIH, clasificándolas según su método de detección.

## 2.2. Estimación de parámetros mediante redes neuronales de Hopfield

Las redes de Hopfield han sido aplicadas a la resolución de problemas de optimización desde poco después de su definición en 1982 [26, 39]. En este trabajo se aplicará esta capacidad a la resolución de problemas de estimación basándose en el hecho de que un problema de estimación puede definirse como un problema de minimización del error de predicción.

Sea un modelo definido en ecuaciones diferenciales ordinarias (EDO) de primer orden  $\frac{dx}{dt} = f(x, \theta)$ , donde  $x$  es el vector de variables de estado y  $\theta$  es el vector de parámetros. A menudo, los parámetros entran de forma lineal en la ecuación del sistema, como es el caso, por ejemplo, en los modelos de sistemas mecánicos rígidos. Asumimos, por tanto, que el modelo puede escribirse como  $\frac{dx}{dt} = A(x)\theta + b(x)$ , de forma que podemos definir el modelo en la forma lineal en los parámetros (LIP):

$$y(x) = A(x)\theta \quad (2.1)$$

donde  $A(x)$  es una matriz y el vector  $y(x)$ , que agrupa los términos que no contienen parámetros, viene dado por  $y(x) = \frac{dx}{dt} - b(x)$ . Hay que tener en cuenta que esta definición implica sólo una leve pérdida de generalidad ya que el sistema es lineal en los parámetros pero no en las variables. De esta forma, el problema de estimación se formulará como el de encontrar los parámetros  $\hat{\theta}$  que minimizan el error de predicción definido como:

$$e = y(x) - A(x)\hat{\theta} \quad (2.2)$$

El método de estimación que se utilizará aquí es una particularización de la red neuronal de Hopfield denominada formulación de Abe [1,

30, 10] que ya ha sido utilizada con éxito en [9, 11]. En general, una red de Hopfield se describe mediante un modelo con tres funciones: la función de activación, la dinámica de evolución y la función de energía o de Lyapunov. En función de como se formule cada uno de esos tres componentes, se obtendrá una particularización de la red. La formulación de Abe utilizada aquí queda definida por el siguiente sistema de ecuaciones:

$$\begin{aligned} s_i(t) &= \tanh\left(\frac{u_i(t)}{\gamma}\right) \\ \frac{du_i}{dt} &= \sum_j w_{ij}s_j - I_i \\ V(s) &= -\frac{1}{2} \sum_i \sum_j w_{ij}s_i s_j + \sum_i I_i s_i \end{aligned} \quad (2.3)$$

donde  $s$  es el estado de la neurona,  $u$  es una variable de estado interna con cierta analogía con el potencial de activación de las neuronas biológicas,  $w$  e  $I$  son los pesos de las conexiones entre las neuronas y  $\gamma$  es un parámetro de diseño que regula la pendiente de la tangente hiperbólica. En esta particularización, la función de activación es la tangente hiperbólica. En general, pueden usarse otras funciones, con la condición de que deben ser funciones continuas, derivables y monótonamente crecientes.

La formulación de Abe puede deducirse de la relación  $\frac{du_i}{dt} = -\frac{\partial V}{\partial s_i}$  que se observa en las dos últimas ecuaciones en (2.3) y que implica que la dinámica de la red evoluciona hacia los mínimos de  $V$ . Así, la optimización con redes neuronales de Hopfield se fundamenta en forzar la equivalencia de la función de energía con la función objetivo, con lo que se conseguirá que la red tienda de forma natural hacia las soluciones del problema de optimización.

Para minimizar el error de estimación se tomará como función objetivo la norma al cuadrado del error de predicción:

$$E = \frac{1}{2} e^\top e = \frac{1}{2} \hat{\theta}^\top A^\top A \hat{\theta} + \hat{\theta}^\top (-A^\top y) + \frac{1}{2} \|y\|^2 \quad (2.4)$$

cuyo último término puede despreciarse ya que no depende de la es-

timación  $\hat{\theta}$ . Si se reescribe la función de energía de la ecuación (2.3) para expresarla de forma matricial:

$$V(s) = -\frac{1}{2}s^\top W s + s^\top I \quad (2.5)$$

se observa claramente que el problema de estimación es definido como un problema de optimización que es resuelto por una red de Hopfield definida por las salidas de las neuronas  $s$ , la matriz de pesos  $W$  y el vector de sesgos  $I$  descritos en la ecuación (2.6):

$$W = -A^\top A; \quad I = -A^\top y; \quad s \equiv \hat{\theta} \quad (2.6)$$

Puesto que los componentes de la matriz  $A$  dependen de los estados variables del sistema, tanto  $W$  como  $I$  son variables en el tiempo.

Un hecho fundamental en este trabajo es que los datos disponibles de la epidemia de VIH/SIDA no tienen una frecuencia suficiente para poder considerar que trabajamos con un sistema continuo. Por ello, el modelo de la epidemia VIH/SIDA en Cuba ha sido definido de forma discreta utilizando la discretización de la ecuación diferencial del modelo continuo mediante la regla de Euler [8]:

$$x_{n+1} = x_n + \Delta t (A(x_n) \theta + b(x_n)) \quad (2.7)$$

donde  $\Delta t$  se elegirá de acuerdo a la disponibilidad de los datos. Reescribiendo el error de predicción como:

$$e_{n+1} = x_{n+1} - x_n - \Delta t (A(x_n) \theta + \Delta t b(x_n)) \quad (2.8)$$

y desarrollando la función objetivo  $E = \frac{1}{2}e^\top e$  para el caso discreto se obtienen los pesos y sesgos de la red:

$$W = -\Delta t^2 A^\top A; \quad I = -\Delta t A^\top (x_{n+1} - x_n) \quad (2.9)$$

con lo que el método quedaría completamente definido para el caso de sistemas discretos.

### 2.3. Modelo de la epidemia de VIH/SIDA en Cuba

En esta sección se presentará el modelo del VIH/SIDA en Cuba que ha sido utilizado en este trabajo. También se definirá un umbral de detección, con un significado análogo al número básico de reproducción, y que permitirá predecir la posible evolución de la epidemia.

El modelo utilizado, definido como un sistema dinámico de poblaciones mediante ecuaciones diferenciales ordinarias, ha sido realizado a partir del conocimiento de expertos en la materia y da una especial relevancia a la influencia de los distintos métodos de detección de personas infectadas. El modelo se describe mediante el sistema de ecuaciones [27]:

$$\begin{aligned}
 \frac{dX}{dt} &= (\lambda - k_1 - \beta - \mu)X + \lambda'(Y_1 + Y_2) - k_2 \frac{X(Y_1 + Y_2)}{X + Y_1 + Y_2} \\
 \frac{dY_1}{dt} &= k_1X - (\mu + \beta')Y_1 \\
 \frac{dY_2}{dt} &= k_2 \frac{X(Y_1 + Y_2)}{X + Y_1 + Y_2} - (\mu + \beta')Y_2 \\
 \frac{dZ}{dt} &= \beta X + \beta'(Y_1 + Y_2) - \mu'Z
 \end{aligned}
 \tag{2.10}$$

donde  $X$  representa la población de individuos infectados de VIH que no han sido detectados por el sistema sanitario,  $Y$  es la población de personas infectadas que sí han sido detectadas y  $Z$  es la población de personas que han desarrollado SIDA. La población de personas infectadas detectadas  $Y$  ha sido subdividida en las poblaciones  $Y_1$ , personas detectadas mediante el programa de notificación de parejas e  $Y_2$ , personas detectadas mediante cualquier otro mecanismo del sistema sanitario y que será denominado como métodos “azarosos”. El objetivo de esto es poder evaluar de forma diferenciada la influencia del programa de notificación de parejas o búsqueda activa, en el que personas que son detectadas como portadores de VIH son instadas a declarar sus últimos contactos sexuales, a los que se les realiza una prueba de seropositividad.

Una forma usual de evaluar la evolución futura de las epidemias

es mediante el número básico de reproducción  $R_0$ , que se define como la media de casos secundarios de la epidemia causados por cada individuo infectado. De esta forma, si  $R_0 > 1$  el número de afectados se incrementará, si  $R_0 < 1$  el número de infectados decrecerá y si  $R_0 = 1$  la epidemia podría ser considerada endémica. La forma más habitual de calcular el valor de  $R_0$  es definirlo como el radio espectral de la matriz jacobiana del sistema en el equilibrio sin enfermedad [19]. Sin embargo, aquí no es posible calcularlo mediante ese procedimiento, ya que la parte derecha de las ecuaciones del modelo ni es diferenciable ni está definida en el origen. Por ello, se calculará una aproximación utilizando las poblaciones normalizadas  $x = \frac{X}{X+Y}$  e  $y = \frac{Y}{X+Y}$ . Si los autovalores de este nuevo sistema son negativos la población de infectados decrecerá, por lo que definiremos un umbral  $D_0 < 1$  equivalente a esta condición, y cuyo significado será, por tanto, análogo a  $R_0 < 1$ . Teniendo en cuenta que la definición del umbral  $D_0$  tiene cierta arbitrariedad, su valor será evaluado sólo cualitativamente, no en su valor cuantitativo, aunque su análisis seguirá siendo muy valioso para conocer la posible evolución de la enfermedad.

Sea la ecuación de la población de infectados desconocidos normalizada,  $\frac{dx}{dt}$ . A partir del modelo definido en 2.10 y la regla de la cadena  $\frac{dx}{dt} = \frac{\partial x}{\partial X} \frac{dX}{dt} + \frac{\partial x}{\partial Y} \frac{dY}{dt}$ , se puede definir:

$$\frac{dx}{dt} = (\lambda' - \lambda + k_2 + \beta - \beta') x^2 + (\lambda - 2\lambda' - k_1 - k_2 - \beta + \beta') x + \lambda' \quad (2.11)$$

Por definición, los valores de los parámetros son siempre positivos, por lo que esta ecuación tendrá una única raíz positiva  $x_0$  que se corresponderá con el punto de equilibrio. Además, por definición,  $y = 1 - x$  de manera que, en particular,  $y_0 = 1 - x_0$ .

Calculando el jacobiano para el sistema descrito en (2.10), obviando la población  $Z$  que este caso no será significativa, y en función de las poblaciones normalizadas, se obtiene:

$$J = \begin{pmatrix} \lambda - k_1 - \beta - \mu - k_2 y^2 & \lambda' - k_2 x^2 \\ k_1 + k_2 y^2 & -(\mu + \beta') + k_2 + x^2 \end{pmatrix} \quad (2.12)$$



cuyo determinante es:

$$|J| = (k_1 + \beta + \mu) (\mu + \beta') - (\mu + \beta') (\lambda + k_2 (x_0 - y_0)) + K_2 (\lambda + \beta' - \beta) x_0^2 - \lambda' (k_2 y_0^2 + k_1) \quad (2.13)$$

en el que se han sustituido los valores de las poblaciones normalizadas por el punto de equilibrio  $(x_0, y_0)$ . Finalmente, calculando una expresión que haga equivalente  $|J| < 0$  y  $D_0 < 1$  se obtiene:

$$D_0 = \frac{\lambda + k_2 (x_0 - y_0)}{k_1 + \beta + \mu} + \frac{\lambda'}{\mu + \beta'} \frac{k_2 y_0^2 + k_1}{k_1 + \beta + \mu} - \frac{\lambda + \beta' - \beta}{\mu + \beta'} \frac{k_2 x_0^2}{k_1 + \beta + \mu} \quad (2.14)$$

Los parámetros de detección  $k_1$  y  $k_2$  son variables en el tiempo, lo que podría llevar a cuestionar este razonamiento, que sólo es válido para parámetros constantes. Sin embargo, se puede considerar que la variación de los parámetros es mucho más lenta que la de la dinámica de la población, por lo que sería válido considerarlos como constantes a trozos. Además, la variación de los parámetros implicará la variación del umbral  $D_0$  y permitirá evaluar los cambios en la evolución de la enfermedad.

## 2.4. Resultados experimentales

Se han realizado una serie de experimentos con el objetivo de validar el comportamiento del método de estimación y de la bondad del propio modelo de la epidemia. Para ello se ha utilizado un registro de 16 años de datos reales de la epidemia de VIH/SIDA en Cuba con información semanal de las poblaciones  $Y$  y  $Z$ . Para evitar los errores debidos a las imprecisiones en el registro, estos datos se han promediado para utilizarlos con una periodicidad cuatrimestral. Los parámetros del modelo distintos de  $k_1$  y  $k_2$  se considerarán constantes y conocidos, ya que su valor puede ser obtenido de manera fiable mediante técnicas estadísticas, y sus valores se resumen en la tabla 2.1. Los datos de la población  $X$ , que son desconocidos pero necesarios para realizar los experimentos, han sido obtenidos integrando una simplificación del modelo de la epidemia en el que se han utilizando los valores de los

Cuadro 2.1: Parámetros considerados conocidos y sus valores.

Tasa	Símbolo	Valor
Contagio causado por individuos no detectados	$\lambda$	0.5750
Contagio causado por individuos detectados	$\lambda'$	0.0403
Desarrollo de SIDA por individuos no detectados	$\beta$	0.1300
Desarrollo de SIDA por individuos detectados	$\beta'$	0.1660
Mortalidad de individuos infectados sanos	$\mu$	0.0110
Mortalidad de pacientes con SIDA	$\mu'$	0.7740

parámetros descritos en la tabla 2.1, más dos valores posibles de  $k_1$  y  $k_2$  obtenidos también con métodos estadísticos.

Los experimentos que han sido seleccionados como los que ofrecen los resultados más interesantes son:

1. Estimación de los parámetros de detección  $k_1$  y  $k_2$ .
2. Cálculo del umbral de detección  $D_0$ , incluyendo una representación de la sensibilidad del modelo respecto al parámetro  $k_2$ .
3. Evaluación del modelo de la epidemia mediante la comparación entre las poblaciones de infectados predichas y reales.

### 2.4.1. Estimación de parámetros de detección

Para la estimación de los parámetros de detección se ha construido una red neuronal de Hopfield de dimensión dos donde los estados de las dos neuronas,  $s_1$  y  $s_2$ , se corresponden con la estimación de los parámetros  $k_1$  y  $k_2$ . Las ecuaciones de las poblaciones  $X$  y  $Z$  no se han incluido en la expresión del error de predicción, utilizado para realizar la estimación de los parámetros, ya que la primera no se corresponde con datos reales y la segunda no influye en los parámetros a estimar.

Los resultados obtenidos se representan en la figura 2.1. Para analizar estos resultados hay que tener en cuenta que el significado de ambos parámetros es distinto, ya que tal como se describía en la sección

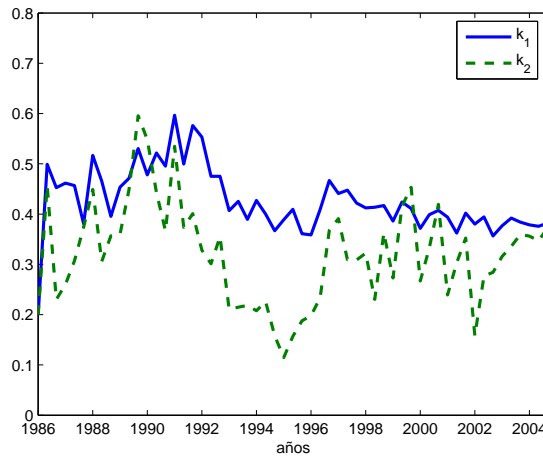


Figura 2.1: Estimación de los parámetros de detección ( $k_1$  y  $k_2$ ) del modelo VIH/SIDA en Cuba.

2.3,  $k_1$  y  $k_2$  ponderan una relación lineal y no lineal respectivamente. Se puede observar que, según los resultados obtenidos, el parámetro  $k_2$  ha pasado por diferentes fases y que son coherentes con las distintas situaciones que ha vivido el programa de notificación de parejas. En la primera fase, hasta 1992, los altos valores del parámetro reflejan la apuesta de las autoridades sanitarias por este tipo de detección. Más tarde, de 1992 a 1996 se puede apreciar una importante caída en los valores de  $k_1$  y  $k_2$  que puede relacionarse con el llamado periodo especial, la crisis económica que vivió la isla tras la caída de la Unión Soviética. A partir de 1997 los valores de ambos parámetros suben, pero  $k_2$  vuelve a bajar, lo que puede relacionarse con la aparición de un nuevo método de detección, la detección gracias a la intervención del médico de familia, método que en este trabajo se ha incluido dentro de los llamados métodos azarosos.

En los valores estimados de los parámetros pueden observarse fuertes oscilaciones que se consideran debidas a la presencia de ruido, que afecta al registro de los datos. Este problema se solventa mediante algoritmos de regularización [20], característica que se ha introducido en el estimador realizando nuevas pruebas en las que se ha limitado

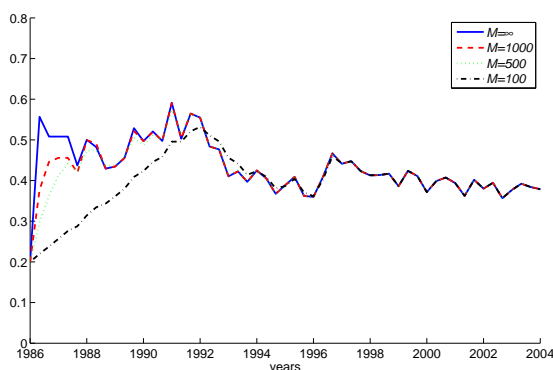


Figura 2.2: Estimación del parámetro  $k_1$  con tiempo final de integración  $t = M h$  ( $h = 10^{-6}$ ).  $M = \infty$  mantiene la integración hasta la convergencia.

el número de pasos de integración, parando la convergencia antes de llegar a un punto estable, y siguiendo un razonamiento similar a los métodos *early stopping* [24]. Como ejemplo de estas pruebas, en la figura 2.2 se representan los resultados obtenidos para el parámetros  $k_1$ . Los resultados nos llevan a pensar que el ajuste del tiempo final de integración tiene un carácter crítico, estableciendo un compromiso entre la velocidad de convergencia y la suavidad del ajuste. En esta línea, es necesario profundizar en el estudio en el futuro. Para el resto de experimentos, descritos a continuación, se considerarán las estimaciones representadas en la figura 2.1.

### 2.4.2. Cálculo de tasa de reproducción de la epidemia

Sustituyendo el valor de los parámetros en la fórmula presentada en la sección 2.3 se ha calculado el valor del umbral de detección  $D_0$  y se ha obtenido el resultado representado en la figura 2.3. Como era de esperar,  $D_0$  pasa por las mismas fases que describimos en la estimación de los parámetros de detección. En la primera fase la incidencia de la epidemia todavía es muy baja, causada fundamentalmente por personas que venían del extranjero y ya habían sido detectadas allí o que llegaban ya en estado sintomático. Más tarde, la epidemia crece

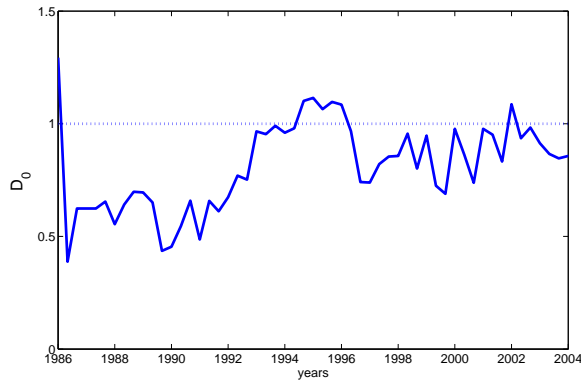


Figura 2.3: Umbral de detección ( $D_0$ ) obtenido para los parámetros estimados, para la epidemia de VIH/SIDA en Cuba.

en el periodo especial y, finalmente, el valor de  $D_0$  se estabiliza cuando el sistema sanitario vuelve a tener más recursos. Hay que tener en cuenta que, según los resultados obtenidos, el descenso en las tasas de detección no implica un descenso en la epidemia, sino un problema en la estrategia sanitaria de detección.

Un ejemplo de la información que se puede extraer con estas pruebas se mostrará analizando la dependencia de  $D_0$  respecto a los parámetros de detección. En este caso, a la vista de los resultados obtenidos puede pensarse que una buena solución para controlar la enfermedad es aumentar los parámetros de detección y conseguir que la epidemia pase al estado de erradicación ( $D_0 < 1$ ). Puesto que la vigilancia del sistema sanitario ya es muy exhaustiva, es muy difícil que  $k_1$  pueda aumentar, por lo que nos centraremos en analizar cuánto ha de aumentar  $k_2$ . En la figura 2.4 se representa el valor de  $D_0$  en función de  $k_2$  para un determinado valor de  $k_1$ , que en este caso ha sido establecido a  $k_1 = 0,38$ , su valor en 1995. Se puede observar que el valor que tendría que haber alcanzado  $k_2$  es mucho mayor del que tenía en esos momentos, un valor que sería inalcanzable para las autoridades sanitarias. En cualquier caso, esto nos da un muestra de que, a pesar de la poca incidencia del VIH/SIDA en Cuba, no es posible conseguir la erradicación de la epidemia a corto plazo.

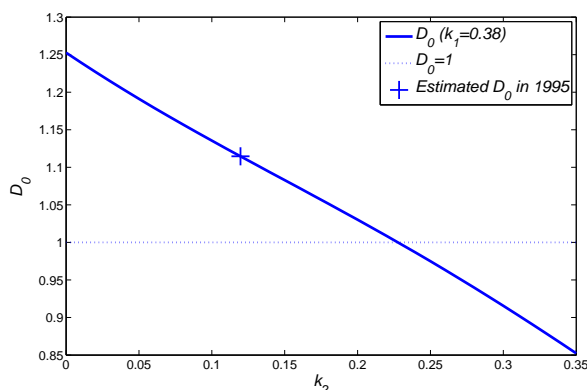


Figura 2.4: Umbral de detección  $D_0$  en función de  $k_2$  cuando  $k_1$  se establece a su valor en 1995. La cruz es el valor de  $D_0$  en 1995 ( $k_2 = 0,12$ ).

### 2.4.3. Predicción de la población infectada

Con el objetivo de evaluar la eficiencia del método de estimación y del modelo de la epidemia, se han calculado las poblaciones predichas de  $Y_1$  e  $Y_2$  para compararlas con las poblaciones reales. Para ello se integrarán las ecuaciones del modelo utilizando los valores estimados de los parámetros y considerando los valores de las poblaciones en 1986 como valores iniciales. Además, se calculará el error cuadrático medio de cada población:

$$MSE = \frac{\sum_{t=t_1}^{t_N} (Y_i(t) - \hat{Y}_i(t))^2}{N} ; \quad i = 1, 2 \quad (2.15)$$

siendo  $N$  el número total de datos. Los resultados obtenidos se representan en las figuras 2.5 y 2.6 y muestran un buen ajuste de las curvas y un error cuadrático medio aceptable, lo que permite afirmar que la estimación es satisfactoria y que el modelo se ajusta bien a la evolución real de la epidemia. Se puede observar un hecho llamativo en la trayectoria de las poblaciones, y es que entre los años 1992 y 1996 las poblaciones permanecen casi constantes. Este hecho es coherente con los resultados anteriores e indicaría el descenso en los trabajos de

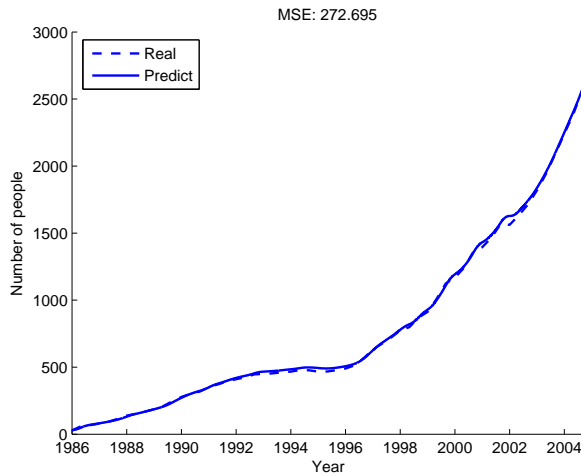


Figura 2.5: Poblaciones real y predicha de los detectados con VIH mediante métodos azarosos en Cuba.

detección debido a la escasez de recursos de ese periodo, no el descenso de la incidencia de la propia enfermedad.

## 2.5. Conclusiones

En este capítulo se ha presentado un método de estimación basado en redes neuronales de Hopfield y se ha aplicado a un modelo de la epidemia de VIH/SIDA en Cuba. En concreto se han estimado los llamados parámetros de detección, presentes en este modelo. La estimación de estos parámetros tiene un especial interés debido a la dificultad de estimarlos mediante métodos clásicos y a su importancia desde el punto de vista epidemiológico. Además, se ha calculado un umbral de detección de la epidemia, denominado  $D_0$ , otro valor clave para estudiar la evolución de la enfermedad debido a que su significado, análogo al del número básico de reproducción  $R_0$ , da una idea sobre el estado de crecimiento o decrecimiento de la epidemia. Los resultados obtenidos para la estimación son coherentes y satisfactorios y se ha concluido que el modelo se ajusta a la epidemia. También se ha analizado la relación entre el umbral y los parámetros de detección y se ha llegado a la conclusión de que, para llevar a la epidemia a un

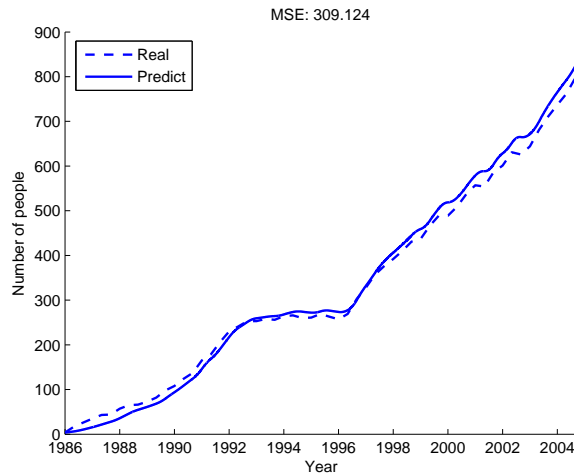


Figura 2.6: Poblaciones real y predicha de los detectados con VIH mediante el programa de notificación de parejas en Cuba.

estado de decrecimiento, habría que aumentar la detección hasta un nivel que no es posible alcanzar al menos corto plazo.

Como líneas futuras a este trabajo se propone aumentar el número de parámetros estimados mediante la red neuronal. Ya se han presentados algunos experimentos preliminares [7] donde se incorporan nuevos parámetros a la estimación. Otra línea de trabajo interesante será la definición de un modelo más complejo donde se incorporen los cambios ocurridos a partir de 2001: la influencia de la detección mediante el médico de familia, los tratamientos con antirretrovirales o la nueva definición de SIDA de la OMS. Otro aspecto importante es conseguir nuevas estimaciones para la población  $X$  de infectados de VIH no detectados por el sistema sanitario. Finalmente, un aspecto interesante desde el punto de vista matemático, es analizar del comportamiento del estimador ante la presencia de ruido.



## Capítulo 3

# Ratio de la infección oculta del VIH en Cuba

### 3.1. Introducción

El objetivo de este capítulo es describir el trabajo correspondiente al artículo “The ratio of hidden HIV infection in Cuba” [6]. En él, se intenta solventar uno de los problemas que se encontraban en el trabajo descrito en el capítulo 2: la falta de información sobre la población  $X$  del modelo del VIH/SIDA en Cuba, que se corresponde con los individuos infectados por VIH que aún no han sido detectados por las autoridades sanitarias. Para ello se definirá el concepto de “ratio de infección oculta” y, a partir de él y del modelo definido en la ecuación (2.10), se propondrá un nuevo modelo para la epidemia de VIH en Cuba. Además, también se realizará el análisis de estabilidad y el análisis de sensibilidad respecto a los parámetros para el nuevo modelo propuesto. Finalmente, se describirán los experimentos realizados, en los que se ha utilizado una base de datos reales de la epidemia en Cuba. El objetivo de estos experimentos será extraer información sobre la epidemia y estudiar la bondad del modelo propuesto.

### 3.2. Modelo de la ratio de infección oculta de VIH en Cuba

En el capítulo 2 se definía un modelo para la enfermedad de VIH/SIDA en Cuba descrito por el sistema de ecuaciones (2.10). Este modelo tiene el inconveniente de que una de sus variables, la población  $X$ , que agrupa los individuos infectados por el VIH que aún no han sido detectados, es desconocida por propia definición. Hay que tener en cuenta que los valores de las otras poblaciones,  $Y$  y  $Z$ , referentes a los seropositivos detectados y a los enfermos de SIDA respectivamente, son conocidos y se registran en bases de datos de la epidemia. Además, el valor de  $X$  es difícil de aproximar porque depende del periodo de incubación de SIDA (parámetro  $\beta$  en el modelo), que tiene un valor alto y bastante variabilidad entre individuos. Obtener información sobre la población  $X$  es crucial por diversas razones, entre ellas, porque permite estudiar la eficiencia de los métodos de control de las autoridades sanitarias en la detección temprana de seropositivos. Puesto que los individuos infectados por VIH son contagiosos durante el resto de su vida, ser desconocedores de su situación puede llevarles a no tomar medidas preventivas y provocar un crecimiento descontrolado de la epidemia. En este sentido, la detección temprana de infectados de VIH constituye, por sí misma, un método de control de la enfermedad.

En este trabajo se propone un nuevo modelo de la epidemia de VIH/SIDA en el que las ecuaciones del modelo no dependen directamente de la población  $X$ , sino de lo que se denomina aquí como *ratio de infección oculta*, definida como el cociente de los individuos infectados de VIH desconocidos entre los conocidos:

$$r = \frac{X}{Y} \quad (3.1)$$

donde  $X$  e  $Y$  son respectivamente las poblaciones de individuos no detectados y de individuos detectados por cualquier método del sistema sanitario, al igual que en el modelo definido en (2.10). Se parte de dicho modelo en el que se ha obviado la ecuación correspondiente a la población  $Z$ , que no afecta en este caso, y se ha agrupado la población

de seropositivos detectados,  $Y = Y_1 + Y_2$ , con lo que queda:

$$\begin{aligned} \frac{dX}{dt} &= (\lambda - k_1 - \beta - \mu) X + \lambda' Y - k_2 \frac{XY}{X + Y} \\ \frac{dY}{dt} &= -(\mu + \beta') Y + k_1 x + k_2 \frac{XY}{X + Y} \end{aligned} \quad (3.2)$$

En este caso, a diferencia del trabajo descrito en el capítulo 2, los parámetros se considerarán constantes en el tiempo. A partir de estas ecuaciones, se puede obtener el modelo dependiente de la ratio  $r$  utilizando la regla de la cadena:

$$\frac{dr}{dt} = \frac{\partial r}{\partial X} \frac{dX}{dt} + \frac{\partial r}{\partial Y} \frac{dY}{dt} \quad (3.3)$$

de donde se obtiene:

$$\frac{dr}{dt} = -k_2 \frac{r(r+1)}{r+1} - k_1 r^2 + (\lambda - k_1 - \beta + \beta') r + \lambda' \quad (3.4)$$

Teniendo en cuenta que las poblaciones  $X$  e  $Y$  son siempre mayores que 0 y que, por tanto,  $r + 1 \neq 0$ . Se obtiene finalmente el siguiente modelo de la epidemia:

$$\frac{dr}{dt} = -k_1 r^2 + a r + \lambda' \triangleq f(r) \quad (3.5)$$

donde para mayor simplicidad se ha definido el parámetro  $a = \lambda - k_1 - k_2 - \beta + \beta'$ . De esta forma, se ha obtenido en (3.5) una expresión de la dinámica del modelo que sólo depende de la propia variable  $r$  y que tiene solución explícita. Resolviendo la ecuación  $f(r) = 0$  se obtiene:

$$\begin{aligned} r_+ &= \frac{-a + \sqrt{a^2 + 4 k_1 \lambda'}}{-2 k_1} \\ r_- &= \frac{-a - \sqrt{a^2 + 4 k_1 \lambda'}}{-2 k_1} \end{aligned} \quad (3.6)$$

Puesto que los valores de todos los parámetros son siempre mayores que cero, el discriminante de la raíz es positivo y, por tanto, las dos raíces de la ecuación son reales, siendo una negativa y la otra positiva. De esta forma, puede decirse que  $f(r) = -k_1(r - r_+)(r - r_-)$  y se llega

a la solución explícita:

$$r(t) = \frac{r_+ - C r_- e^{-k_1(r_+ - r_-)t}}{1 - C e^{-k_1(r_+ - r_-)t}} \quad (3.7)$$

donde  $C > 0$  es la constante de integración.

### 3.2.1. Análisis de estabilidad

Las dos raíces que se han obtenido para la ecuación (3.5),  $r_-$  y  $r_+$ , serán los puntos fijos del sistema. De ellas, sólo  $r_+$  es un valor posible, ya que  $r > 0$ . A partir de la solución explícita descrita en 3.7 la tendencia de las trayectorias a largo plazo será:

$$\lim_{x \rightarrow \infty} r(t) = r_+ \quad (3.8)$$

por lo que se puede concluir que el punto fijo  $r_+$  es globalmente asintóticamente estable para  $r > 0$ .

Puesto que el modelo propuesto dispone de solución explícita, no sería necesario realizar un análisis de estabilidad convencional. Sin embargo, merece la pena realizarlo ya que futuras modificaciones del modelo podrían dar lugar a un sistema de EDO sin solución. Un ejemplo de ello es el término no lineal relacionado con la búsqueda activa,  $\frac{xy}{x+y}$ , presente en el modelo (2.10) y que está todavía sujeto a discusión.

Para analizar la estabilidad del modelo se estudiarán tres conceptos [25, 31]: los puntos fijos, la estabilidad asintótica de los puntos fijos y la cuenca de atracción.

En primer lugar, se evaluará su jacobiano en el punto fijo  $r_+$ , que en este caso se corresponde con la derivada:

$$f'(r) = -2 k_1 r_+ + a = -\sqrt{a^2 + 4 k_1 \lambda'} < 0 \quad (3.9)$$

Esta expresión es siempre negativa, por lo que el punto fijo  $r_+$  es asintóticamente estable. Una vez probada la estabilidad del punto fijo hay que estudiar la cuenca de atracción, es decir, los valores del sistema que tienden al punto fijo. Para ello se definirá una función candidata de Lyapunov,  $V(r)$  [41], definida de forma que cumpla  $\frac{dr}{dt} = -\frac{\partial V}{\partial r}$ , con

el objetivo de obtener en todo caso  $\frac{dV}{dt} \leq 0$ , puesto que  $\frac{\partial V}{\partial r} \frac{dr}{dt} = -\frac{dr}{dt} \frac{dr}{dt}$ .

Defínase una función  $V_0(r)$  como:

$$V_0(r) = -\left(-\frac{k_1}{3}r^3 + \frac{a}{2}r^2 + \lambda' r\right) \quad (3.10)$$

que tiene las siguientes tres raíces:

$$V_0(r) = 0 \implies r = \begin{cases} 0 \\ \frac{3a + \sqrt{9a^2 + 48k_1\lambda'}}{4k_1} \\ \frac{3a - \sqrt{9a^2 + 48k_1\lambda'}}{4k_1} \end{cases} \quad (3.11)$$

Las tres raíces de  $V_0(r)$  son reales y sólo la segunda es positiva, que además es mayor que el punto fijo  $r_+$ . Considerando las raíces obtenidas para  $V_0(r)$  y el carácter cúbico de la función, se puede definir entonces:

$$V(r) = -\left(-\frac{k_1}{3}r^3 + \frac{a}{2}r^2 + \lambda' r\right) + \min(V_0(r)) \quad (3.12)$$

para asegurar que la función  $V(r)$  es positiva a partir de su primera raíz. Hay que tener en cuenta además que, puesto que  $V_0(r)$  tiene una raíz negativa,  $V(r)$  tendrá también una raíz negativa y que será menor que la raíz de  $V_0(r)$ . En la figura 3.1 se puede observar una representación gráfica de este razonamiento.

Como se cumplen las condiciones para que  $V$  sea una Función de Lyapunov en  $(r \rightarrow \infty)$ , esto es,  $V(r) > 0$  al menos para  $r \geq r_-, r \neq r_+$  y  $\frac{dV}{dt} < 0 \forall r(t) \neq r_-, r_+$ , se puede concluir que el punto fijo  $r_+$  es globalmente asintóticamente estable para los valores positivos de  $r$ . Por tanto, como  $r > 0$ , para cualquier valor inicial el valor de  $r$  siempre converge a  $r_+$ .

### 3.2.2. Análisis de sensibilidad

De forma general, el análisis de sensibilidad de un modelo paramétrico permite estudiar cómo cambia su dinámica respecto a cambios en cada uno de los parámetros que contiene [17]. El interés se centrará en

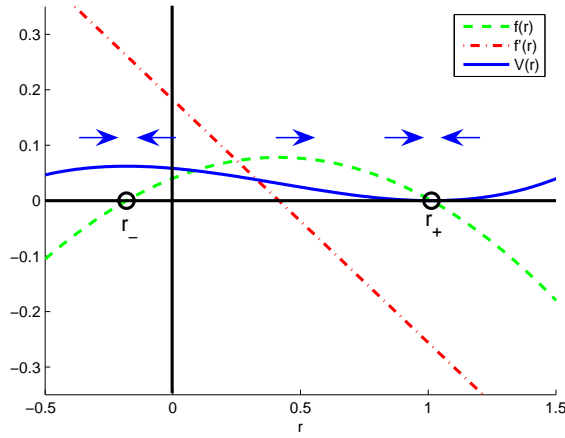


Figura 3.1: Función de Lyapunov,  $V(t)$ . Las flechas muestran las trayectorias del sistema hacia el punto fijo  $r_+$ . Obsérvese que la función  $f'(r)$  es positiva para el punto fijo inestable de  $r_-$  y negativa para el estable  $r_+$ .

las derivadas parciales del modelo, por lo que cada parámetro puede ser tratado uno a uno manteniendo el resto de parámetros fijos. Para cada parámetro  $\theta \in [\lambda, \lambda', \beta, \beta', k_1, k_2]$  puede definirse su ecuación de sensibilidad como:

$$\psi_\theta = \frac{\partial r}{\partial \theta} \tag{3.13}$$

Bajo condiciones suficientes de suavidad, la ecuación de la sensibilidad se puede definir como una EDO:

$$\frac{d\psi_\theta}{dt} = \frac{d}{dt} \frac{\partial r}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{dr}{dt} = \frac{\partial f}{\partial r} \frac{\partial r}{\partial \theta} + \frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial r} \psi_\theta + \frac{\partial f}{\partial \theta} \tag{3.14}$$

En el caso de este modelo, se obtiene  $\frac{\partial f}{\partial r} = -2k_1 r + a$ , y las derivadas parciales respecto al vector de parámetros,  $\frac{\partial f}{\partial \theta}$ , son:

$$\frac{\partial f}{\partial \lambda} = \frac{\partial f}{\partial \beta'} = r \frac{\partial f}{\partial \lambda'} = 1 \frac{\partial f}{\partial \beta} = \frac{\partial f}{\partial k_2} = -r \frac{\partial f}{\partial k_1} = -r(r+1) \tag{3.15}$$

De esta forma, para cada parámetro se obtiene un sistema de EDO de dos ecuaciones:

$$\begin{aligned} \frac{dr}{dt} &= -k_1 r^2 + ar + \lambda' \\ \frac{d\psi_\theta}{dt} &= (-2k_1 r + a)\psi_\theta + \frac{\partial f}{\partial \theta} \end{aligned} \quad \theta \in \forall \{\lambda, \lambda', \beta, \beta', k_1, k_2\} \quad (3.16)$$

La primera ecuación está desacoplada de la segunda ya que la sensibilidad no depende de los estados del sistema. Por ello, a la hora de resolver el sistema de ecuaciones, no habrá que calcular la solución numérica para cada parámetro, sino que basta con sustituir en la ecuación de la sensibilidad la solución explícita obtenida en (3.7) e integrar dicha ecuación de sensibilidad para cada caso.

Respecto a la estabilidad, el jacobiano de este sistema es:

$$J = \begin{pmatrix} f'(r) & 0 \\ D & f'(r) \end{pmatrix} \quad (3.17)$$

que tiene un único autovalor de multiplicidad dos y que nos lleva al jacobiano del sistema original, por lo que el sistema hereda su punto fijo,  $r_+$ , definido en 3.6. Así, la sensibilidad respecto a cada parámetro  $\psi_\theta$  tiende a:

$$\lim_{t \rightarrow \infty} \psi_\theta = - \frac{\frac{\partial f}{\partial \theta}}{\frac{\partial f}{\partial r}} \Big|_{r=r_+} = - \frac{\frac{\partial f}{\partial \theta} \Big|_{r=r_+}}{-\sqrt{a^2 + 4k_1 + \lambda'}} \quad (3.18)$$

Como se ha mencionado, las ecuaciones de la sensibilidad pueden tratarse por separado para cada parámetro, por lo que el proceso para hallar una solución numérica consistirá en: integrar la ecuación del sistema para hallar la trayectoria de  $r$ , evaluar el jacobiano y la derivada  $\frac{\partial f}{\partial \theta}$  en cada instante de tiempo, y resolver las ecuaciones desacopladas. En la figura 3.2.2 se representan las ecuaciones de sensibilidad obtenidas mediante este procedimiento, utilizando los valores de los parámetros que aparecen en la tabla 2.1 y un punto inicial  $r(0) = 500$ . La elección de este valor es bastante arbitraria, ya que no se tienen

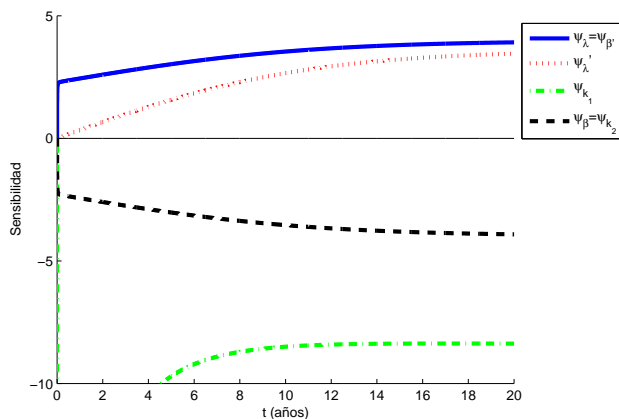


Figura 3.2: Ecuaciones de sensibilidad del modelo de la ratio de infección oculta.

datos fiables sobre él, pero no es algo crucial ya que el comportamiento es muy similar para un amplio rango de valores iniciales.

En los resultados obtenidos se observa que las trayectorias de la sensibilidad convergen a un punto fijo. El término al que el sistema es más sensible es a  $k_1$ , es decir, los cambios en este parámetro son los cambios que más influirán en la dinámica del sistema. También se puede observar que la sensibilidad en todos los casos tiene un régimen transitorio estabilizándose con el tiempo en un valor determinado. Además el módulo de la sensibilidad del sistema es creciente para todos los parámetros, es decir, la dependencia va aumentando con el tiempo en todos los casos excepto en el de  $k_1$ , que tiene un pico máximo en el régimen transitorio. Desde el punto de vista epidemiológico este resultado es coherente, ya que al principio de declararse la enfermedad, cuando  $r = \frac{X}{Y}$  es cercano a infinito, incluso un pequeño número de detectados por los métodos azarosos (parámetro  $k_1$ ) contribuye a que  $r$  decrezca notablemente, mientras que la detección por la búsqueda activa (parámetro  $k_2$ ) es todavía despreciable debido a que la población  $y$  de infectados no detectados tiene un valor pequeño.



Cuadro 3.1: Valores para los parámetros de detección.

Descripción	Símbolo	Valor
Tasa de detección no relacionada con la búsqueda activa	$k_1$	0.5750
Tasa de detección relacionada con la búsqueda activa	$k_2$	0.0403

### 3.3. Pruebas con datos reales

El objetivo de este apartado es validar el modelo propuesto para la ratio de infección oculta utilizando datos reales, a la vez que se extrae información sobre la epidemia. Los datos disponibles corresponden a la misma base de datos que se utilizó en el capítulo 2 ampliada en dos años, es decir, un registro de 22 años (entre 1986 y 2008) con la población de personas que han desarrollado SIDA y las distintas poblaciones de personas infectadas de VIH. En este caso, todos los parámetros en el modelo se considerarán constantes en el tiempo. Sus valores se corresponden con los que aparecen en la tabla 2.1 junto con los valores de  $k_1$  y  $k_2$  en la tabla 3.1, que han sido obtenidos también mediante métodos estadísticos.

Si se reescribe la población de los seropositivos desconocidos sustituyendo  $X = rY$ , se obtiene la EDO lineal:

$$\frac{dY}{dt} = \left( -(\mu + \beta') + k_1 r + k_2 \frac{r}{r+1} \right) Y \quad (3.19)$$

Esta ecuación no puede resolverse directamente, pero teniendo en cuenta que el sistema tiende en todo el dominio al punto fijo  $r_+$ , como se concluyó en el apartado 3.2.1, se puede suponer que a partir de un instante de tiempo en el que el sistema haya alcanzado un régimen estacionario el valor de  $r$  es constante. De esta forma la ecuación de  $Y$  es lineal con respecto a  $r$  y su solución es:

$$Y = y_0 e^{b(t-t_0)} \quad (3.20)$$

donde  $b = -(\mu + \beta') + k_1 r + k_2 \frac{r}{r+1}$ .

Utilizando esta expresión se ha realizado un ajuste lineal a los datos reales de la población  $Y$ . Para realizarlo se han quitado los 4 primeros años de la epidemia (se comienza en 1990), con el objetivo de permitir que el sistema haya llegado a un punto estacionario. Además, se ha realizado un ajuste en dos partes (antes y después del año 2000) con el doble objetivo de, por un lado, realizar un buen ajuste del punto inicial y de la parte final de la curva de  $Y$  y, por otro, tener en cuenta los cambios ocurridos en el año 2000 tanto en la definición de la enfermedad como en el cambio de las circunstancias de la detección. Por tanto, se han considerado dos periodos distintos, [1990,2000] y [2000,2008] ajustando los valores de  $y_0$  y  $b_0$  del modelo exponencial para cada uno de ellos.

Los resultados de la regresión exponencial se representan en la figura 3.3 y su valor numérico se recoge en la tabla 3.2. En ella, se incluyen los valores de los coeficientes reales  $y_0$  y  $b_0$  y estimados  $\hat{y}_0$  y  $\hat{b}_0$ , los valores del Error Absoluto Medio (EAM) y Error Relativo Medio (ERM) del ajuste, definidos como se describe en la ecuación 3.21 (donde  $N$  es la dimensión de los datos), y el coeficiente de determinación  $R^2$ . Se puede observar que este último está bastante por encima de 0.85, que es considerado el valor mínimo a partir del que se considera un ajuste satisfactorio.

$$\begin{aligned}
 MAE &= \frac{1}{N} \sum_{i=1}^N |y(i) - \hat{y}(i)| \\
 MRE &= \frac{1}{N} \sum_{i=1}^N \left| \frac{y(i) - \hat{y}(i)}{y(i)} \right|
 \end{aligned}
 \tag{3.21}$$

En la tabla 3.2 puede observarse que los resultados son significativamente mejores para el segundo periodo de tiempo. Esto no se debe a que en el primer periodo no se ha alcanzado el equilibrio, ya que los errores no se concentran al principio del mismo. Una de las razones de que existan estos errores pueden ser las duras condiciones económicas que había en Cuba en los años 90, que dificultaban las tareas de registro y pudieron provocar errores en los datos. También es relevante el hecho de considerar constantes en el tiempo los parámetros

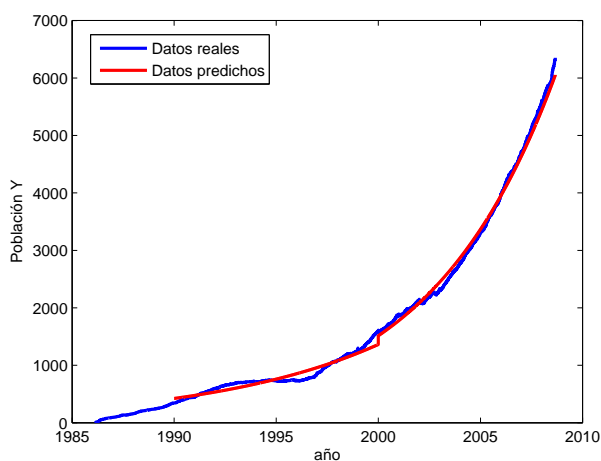


Figura 3.3: Regresión exponencial de la población  $y$  para los periodos [1990,2000] y [2000,2008]

Cuadro 3.2: Coeficientes y errores obtenidos para  $y$  en cada periodo. Error Absoluto Medio (EAM), Error Relativo Medio (ERM) y coeficiente de determinación ( $R^2$ )

Descripción	Periodo 1	Periodo 2	Datos Completos
$t_0$	1990	2008	
$y_0$	346	1598	
$\hat{y}_0$	430	1533	
$b$	0.1636	0.1636	0.1636
$\hat{b}$	0.1165	0.1600	
MAE	65.9584	81.3807	73.0843
MRE	0.0846	0.0278	0.0584
$R^2$	0.9234	0.9954	0.9969

de detección  $k_1$  y  $k_2$ , que son muy dependientes de los recursos de los que disponen las autoridades sanitarias en cada momento. Por otro lado, otra posible razón de los errores en el ajuste es el propio modelo del VIH/SIDA en Cuba, que se puede considerar todavía sujeto a discusión.

### 3.4. Conclusiones

En este capítulo se ha definido un nuevo concepto, la ratio de infección oculta,  $r$ , que nos ha permitido obtener información sobre la población de infectados de VIH que no han sido detectados. Partiendo de esta definición y del modelo para la epidemia de VIH/SIDA en Cuba descrito en el capítulo 2, se ha definido un modelo en el que  $r$  aparece como variable. Para ese modelo se ha realizado el análisis de estabilidad en el que se ha obtenido que el sistema es globalmente asintóticamente estable para los valores positivos de  $r$ , que son, por otro lado, los únicos valores que tienen sentido físico ya que  $r$  es un cociente de poblaciones. También se ha realizado el análisis de sensibilidad del sistema en el que se ha obtenido que el parámetro para el que es más sensible es  $k_1$ , el relacionado con la detección mediante métodos azarosos. Finalmente, se ha validado el modelo con una base de datos reales de la epidemia y el resultado obtenido ha sido satisfactorio.

Uno de los resultados que puede extraerse a partir del trabajo realizado es el valor de la ratio de infección oculta en el punto fijo para los parámetros descritos en las tablas 2.1 y 3.1, esto es  $r_+^* = 1,1$ . En trabajos previos [2, 28] se hablaba de que el número de personas detectadas respecto al global de seropositivos estaría en torno al 75 %, mientras que según el valor de  $r_+^*$  obtenido aquí estaría por debajo del 50 %. La razón principal de esta diferencia es que en estos artículos la búsqueda activa de contactos es un término lineal  $k_2X$ . Si consideramos el término no lineal como un término lineal multiplicado por una tasa variable,  $\frac{Y}{X+Y} = \frac{1}{1+r}$ , la estimación con el término lineal asumiría  $\frac{1}{1+r} \approx 1$ , es decir parte de la suposición  $r \ll 1$ . Es probable que esta suposición no pueda asumirse en este caso, sobre todo a la

luz de los buenos resultados obtenidos en este trabajo para el ajuste de los datos reales. En cualquier caso, el modelado de la búsqueda activa de contactos en el modelo del VIH/SIDA sigue siendo un tema abierto. Una interesante línea de trabajo futuro en este sentido puede ser reemplazar el término no lineal por una función genérica  $g(r)$  en la que se mantengan las condiciones que permiten considerar válido el análisis de estabilidad realizado en la sección 3.2.1, esto es,  $g(r) > 0$ ,  $g(0) = 1$  y  $g'(r) < 0$ . Otra línea de trabajo futura respecto al refinamiento del modelo es incluir distintas condiciones que no se tienen en cuenta en este caso, como son las circunstancias socio-económicas o la aparición de otros métodos de detección tales como el médico de familia. Por otro lado, también sería interesante la obtención de valores más fiables para los parámetros del modelo, añadiendo la posibilidad, además, de que dichos parámetros sean variables en el tiempo.



## Capítulo 4

# Identificación de un modelo de dengue en Cuba

### 4.1. Introducción

En este capítulo se describe el trabajo publicado en el artículo “Hopfield networks for identification of delay differential equations with an application to dengue fever epidemics in Cuba” [23]. En dicho trabajo, el método de estimación de parámetros basado en redes neuronales de Hopfield descrito en la sección 2.2, se aplica a la identificación de un sistema de ecuaciones diferenciales con retardo (EDR). Concretamente, el sistema de ecuaciones sobre el que se aplicará el método de estimación es un nuevo modelo para la epidemia de dengue en Cuba cuya definición forma parte también de las contribuciones del trabajo. Las principales particularidades del modelo propuesto son dos: está definido en ecuaciones diferenciales con retardo, y posee parámetros variables en el tiempo. Desde el punto de vista epidemiológico, estas dos características suponen una ventaja considerable, al evitar la utilización de datos sobre los mosquitos transmisores de dengue, información que es muy difícil de obtener. Desde el punto de vista matemático, la aplicación del estimador basado en redes neuronales de Hopfield a un sistema de ecuaciones con retardo es una interesante novedad.

## 4.2. Modelo de la fiebre de dengue en Cuba

### 4.2.1. La fiebre de dengue

La fiebre de dengue es una enfermedad causada por un virus, de sintomatología similar a una gripe severa. La enfermedad es transmitida mediante la picadura de mosquito, por lo que forma parte de las denominadas enfermedades transmitidas por vectores, es decir, enfermedades que son transmitidas a través de un agente o mecanismo infeccioso desde las personas infectadas a las personas sanas. Más detalladamente, la dinámica de transmisión del dengue es la siguiente: cuando un mosquito pica a una persona infectada, el mosquito pasa por un periodo de incubación del virus de varios días, y a partir de entonces, y durante el resto de su vida, puede transmitir la enfermedad. Por otro lado, cuando un humano se infecta de dengue, resulta aquejado de la enfermedad durante varios días en los que puede infectar a nuevos mosquitos. Cuando, una vez transcurrido ese tiempo, el humano se recupera, no puede infectarse de nuevo con el mismo virus. Ahora bien, dado que se conocen 4 cepas distintas del virus, hay que señalar que la infección por dengue produce inmunidad frente a la cepa del virus por la que se ha sido infectado, pero sólo inmunidad transitoria y parcial frente a las otras cepas. Sin embargo, puesto que aquí se estudiará la enfermedad en Cuba, donde la epidemia no es endémica, y el modelo se definirá para un periodo de tiempo corto, esto es, para un único brote, se podrá considerar que los individuos recuperados no volverán a padecer la enfermedad.

### 4.2.2. Definición del modelo de dengue en Cuba

Existen diversos estudios interesantes sobre modelado e identificación de la epidemiología de dengue [21, 16]. Sin embargo, el principal problema que comparten estos modelos es que incluyen entre sus ecuaciones las poblaciones de mosquitos. Esto supone una limitación severa para trabajar con casos reales de la epidemia, ya que es prácticamente imposible tener datos estadísticos fiables de las poblaciones de mosquitos, especialmente cuando se necesitan datos diferenciados de las



poblaciones de los mosquitos infecciosos y los no infecciosos. Con la idea de evitar este problema, en este capítulo se presenta un modelo en el que se evita incluir de forma explícita dichas poblaciones, es decir, las ecuaciones del modelo describen sólo a poblaciones humanas. El modelo se define como un modelo SIR, en el que las poblaciones se dividen en individuos susceptibles (S), infectados (I) y recuperados (R), y en el que se considera que los individuos recuperados no pueden volver a padecer la enfermedad.

El modelo propuesto se define mediante el sistema de ecuaciones diferenciales con retardo mostrado en la ecuación (4.1):

$$\begin{aligned} \frac{dS}{dt} &= -\beta(t) \frac{S(t)}{N} \sum_{\tau=t_1}^{\tau=t_2} I(t-\tau) e^{-\mu_M \tau} + b N(t) - \mu_H S(t) \\ \frac{dI}{dt} &= \beta(t) \frac{S(t)}{N} \sum_{\tau=t_1}^{\tau=t_2} I(t-\tau) e^{-\mu_M \tau} - r I(t) - \mu_H I(t) \\ \frac{dR}{dt} &= r I(t) - \mu_H R(t) \end{aligned} \quad (4.1)$$

donde el parámetro  $\beta(t)$  es la tasa de infección, el parámetro  $\mu_H$  es la tasa de mortalidad humana,  $\mu_M$  es la tasa de mortalidad del mosquito,  $b$  es la tasa de nacimiento,  $r$  es la tasa de recuperación de la enfermedad y  $N = S + I + R$  es la población total.

En este modelo, para evitar incluir de forma explícita las poblaciones de mosquitos, se considera que las infecciones de nuevos humanos son causadas directamente por la población de humanos infectados. Puesto que en la dinámica de transmisión de dengue un mosquito infectado será un agente infeccioso a partir de que haya pasado su periodo de incubación, podría decirse que la infección de nuevos individuos está causada por el “contacto” con individuos infectados en el pasado. La medida de este contacto, relacionada con la eficiencia de contagio del mosquito en su ciclo de vida, es lo que se ha representado en el modelo con el parámetro  $\beta(t)$ . Este parámetro tiene un interpretación compleja y no puede medirse mediante métodos estadísticos, a diferencia del resto de parámetros del modelo. Por ello, en este trabajo se propone estimar  $\beta(t)$  utilizando el método basado en redes neuro-

nales de Hopfield que fue descrito en la sección 2.2 para la estimación de parámetros en sistemas de ecuaciones diferenciales ordinarias, pero adaptándolo para que pueda ser utilizado con sistemas de ecuaciones diferenciales con retardo.

El modelo propuesto se basa en un razonamiento intuitivo que puede analizarse en cada una de sus ecuaciones de la siguiente manera:

- S: individuos susceptibles. El crecimiento de esta población viene dado por los nacimientos,  $bN$ , mientras que el descenso se debe a los fallecimientos,  $-\mu_H S$ , y a las nuevas infecciones,  $-\beta(t) \frac{S(t)}{N} \sum_{\tau=t_1}^{\tau=t_2} I(t-\tau) e^{-\mu_M \tau}$ .
- I: individuos infectados. Esta población crece por las nuevas infecciones,  $-\beta(t) \frac{S(t)}{N} \sum_{\tau=t_1}^{\tau=t_2} I(t-\tau) e^{-\mu_M \tau}$ , y los nacimientos  $bN$ . La disminución se debe a la recuperación de los afectados,  $-rI$ , y a la mortalidad,  $-\mu_H I$ .
- R: individuos recuperados. Esta población crece por las nuevas recuperaciones,  $rI$ , y los nacimientos  $bN$ , y decrece por la mortalidad de los individuos recuperados,  $-\mu_H R$ .

En el sumatorio que modela las nuevas infecciones, se toma como primer límite del retardo el periodo de incubación del mosquito, 8 días, ya que es el tiempo mínimo que necesita un mosquito para ser infeccioso. El límite superior del sumatorio es el tiempo medio de vida del mosquito, 15 días. Además, cada componente del sumatorio entre 8 y 15 días está ponderada por un término exponencial que ajusta la tasa de supervivencia del mosquito. Por tanto, cada término del sumatorio es la fracción de los mosquitos infecciosos surgidos por los humanos infectados  $\tau$  días antes. Además, el número de nuevas infecciones en  $t$  depende de la probabilidad de que un mosquito infectado pique a un humano susceptible  $\frac{S(t)}{N}$ . Por otro lado, se consigue que la población de mosquitos aparezca sólo de forma implícita utilizando la dependencia temporal de  $\beta(t)$  y los términos con retardo.

### 4.2.3. Número básico de reproducción

Se ha definido el número básico de reproducción,  $R_0$ , para el modelo de dengue propuesto, tal como se hizo para el modelo de VIH/SIDA en la sección 2.3.

En este caso, tampoco se puede utilizar la definición de  $R_0$  como el radio espectral de la matriz jacobiana del sistema en el equilibrio sin enfermedad, así que se calculará también una aproximación. Se partirá de la ecuación de individuos infectados del modelo definido en la ecuación (4.1). Se considerará que la población de individuos susceptibles,  $S$ , es constante para un periodo de tiempo “corto” y que la variación de la población infectada entre los periodos de tiempo  $t$  y  $t - \tau$  es despreciable, es decir, la población está en un punto de equilibrio  $I(t) = I(t - \tau)$  para todo  $\tau = t_1, \dots, t_2$ . Con estas premisas, la ecuación de individuos infectados queda como:

$$\frac{dI}{dt} = \left( \beta(t) \frac{S}{N} \sum_{\tau=t_1}^{\tau=t_2} e^{-\mu_M \tau} - r - \mu_H \right) I(t) \quad (4.2)$$

A partir de este punto de equilibrio, un avance de un periodo de tiempo “corto” provocará un aumento de la población de infectados siempre y cuando se cumpla que:

$$\beta(t) \frac{S}{N} \sum_{\tau=t_1}^{\tau=t_2} e^{-\mu_M \tau} - r - \mu_H > 0 \quad (4.3)$$

y se producirá un descenso si dicho término es negativo. Si se reorganiza la ecuación 4.3, se puede concluir que una aproximación al número básico de reproducción puede ser:

$$R_0 = \frac{\beta(t) \frac{S}{N} \sum_{\tau=t_1}^{\tau=t_2} e^{-\mu_M \tau}}{r + \mu_H} \quad (4.4)$$

Hay que tener en cuenta una serie de premisas para la definición anterior, especialmente si la comparamos con la definición clásica del  $R_0$  para sistemas lineales. Por un lado, mientras que la reorganización de la ecuación 4.3 tiene sentido desde un punto de vista epidemiológico,

desde un punto de vista algebraico es una reordenación arbitraria. Por tanto, la definición de  $R_0$  aportada no es única ni puede ser comparable con otras epidemias ni con otras particularizaciones del dengue: su valor ofrecerá información cualitativa del estado en que se encuentra la epidemia, esto es, expansión, erradicación o equilibrio, pero no tiene sentido como valor absoluto.

### 4.3. Estimador de Hopfield en sistemas con retardo

En la sección 2.2 se describe el método de estimación basado en redes neuronales de Hopfield que será utilizado en este capítulo, pero la definición de dicha sección se limitaba a modelos dados por ecuaciones diferenciales ordinarias. Para aplicarlo al caso de dengue, es necesario extender la definición para el caso de ecuaciones diferenciales con retardo. Retomando la descripción para el caso discreto, donde el sistema se definía como:

$$x_{n+1} = x_n + \Delta t A(x_n) (\theta_n + \theta) \quad (4.5)$$

solamente quedaría generalizarlo para que se adapte a un sistema con retardos, que puede definirse como:

$$x_{n+1} = x_n + \Delta t A(x_n, x_{n-i_1}, x_{n-i_2}, \dots) (\theta_n + \theta) \quad (4.6)$$

Paradójicamente, la discretización del modelo ayuda a definir el sistema con retardos, ya que la generalización para este caso consistiría simplemente en la generalización de la matriz  $A$  para que dependa de estados pasados, donde  $i_1, i_2$ , etc. son los retardos. De esta forma, el estimador queda completamente definido para el modelo epidemiológico con retardo, que constituye un sistema de ecuaciones discretas con retardo desde el punto de vista matemático, simplemente utilizando la generalización de la matriz  $A$  en los pesos y sesgos de la red de

Cuadro 4.1: Parámetros y valores característicos de la enfermedad de dengue.

Descripción	Símbolo	Valor
Mortalidad del mosquito	$\mu_M, t_1^{-1}$	$\frac{1}{15}$
Tasa de recuperación de dengue	$r$	$\frac{1}{7}$
Mortalidad humana	$\mu_H$	$2,4 \cdot 10^{-5}$
Natalidad (considerada igual a la mortalidad)	$b$	$2,4 \cdot 10^{-5}$
Población total (considerada constante)	$N$	$2,18 \cdot 10^6$
Periodo de incubación de dengue en el mosquito	$t_2$	8

Hopfield:

$$W = -\Delta t^2 A^\top A; \quad I = -\Delta t A^\top (x_{+1} - x_n) \quad (4.7)$$

y, por tanto, en la dinámica de la red:

$$\frac{du_i}{dt} = \sum_j w_{ij} s_j - I_i \quad (4.8)$$

#### 4.4. Pruebas con datos reales

En esta sección se describen las pruebas realizadas para evaluar la bondad del estimador neuronal y del modelo de dengue propuesto. Para realizar estas pruebas se ha utilizado una base de datos reales de la epidemia en la que se registra el número de personas infectadas durante los 274 días del brote de dengue en La Habana en 2001. A excepción del parámetro  $\beta$ , todos los parámetros presentes en el modelo propuesto de la epidemia se considerarán conocidos, ya que pueden ser calculados de manera fiable mediante métodos estadísticos. Sus valores han sido fijados a los descritos en la tabla 4.1.

A partir de la población de infectados disponible en la base de datos y de los parámetros conocidos de la epidemia, se han calculado las poblaciones de individuos susceptibles y recuperados. Utilizando

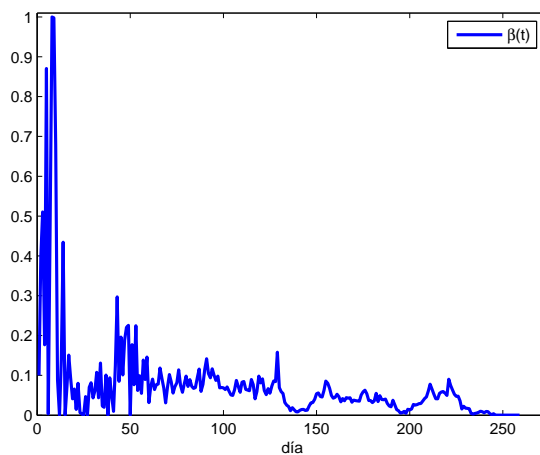


Figura 4.1: Estimación del parámetro  $\beta$  para el brote de dengue de La Habana en 2001

estas poblaciones, se ha construido una red neuronal de dimensión uno con la que se ha estimado el valor del parámetro  $\beta$  y cuyo resultado se representa en la figura 4.1

Con los valores obtenidos para el parámetro  $\beta$  se han realizado dos pruebas más. En la primera de ellas se ha calculado la población predicha de infectados de dengue integrando las ecuaciones del modelo. Los valores iniciales de las poblaciones son los correspondientes al instante en el que la población de infectados ha alcanzado un 10 % de la media de la serie de datos. En la figura 4.2 se representa el valor de la población predicha junto con la población registrada original de infectados. Finalmente, el resultado de  $\beta$  obtenido ha sido utilizado también para el cálculo del número básico de reproducción  $R_0$ , representado en la figura 4.3.

Como puede observarse, los resultados obtenidos son coherentes entre sí y con la definición del modelo. La estimación del parámetro  $\beta$  nos permite extraer información de la población de mosquitos, ya que está fuertemente relacionado con ella. Los valles que aparecen en su gráfica, como los que se aprecian en torno a los días 160 y 220, pueden estar relacionados con un fuerte descenso de la población de mosquitos debido a una campaña de fumigación exitosa. Los valores

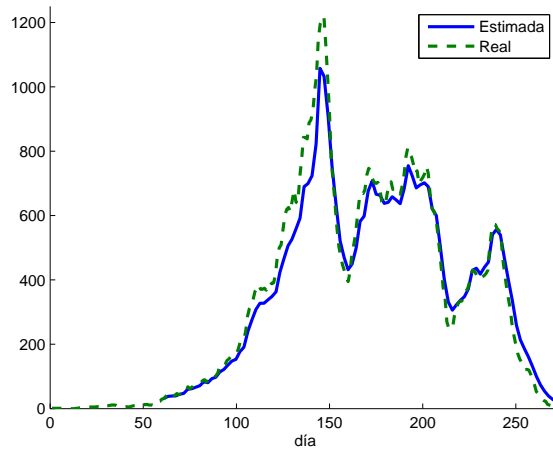


Figura 4.2: Poblaciones real y estimada de individuos infectados de dengue en el brote de La Habana en 2001

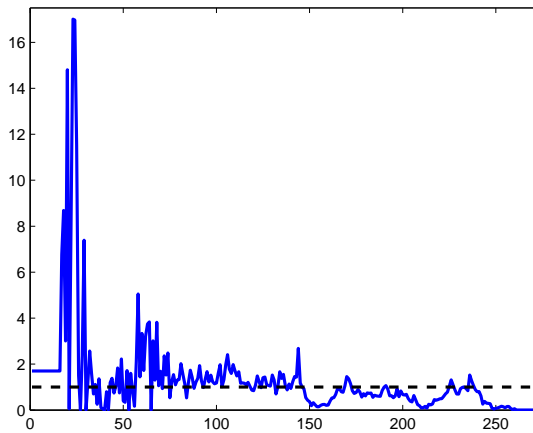


Figura 4.3: Número básico de reproducción  $R_0$  para el brote de La Habana en 2001

de  $\beta$  son muy variables debido probablemente a la presencia de imprecisiones, o ruido, en los datos. En cualquier caso, la interpretación de  $\beta$  debe ser tomada más como la de una tendencia que como la de un valor preciso. Por otro lado, los resultados representados en la figura 4.2 son satisfactorios y nos muestran que hay un buen ajuste de la población predicha a la población real de infectados, y a la espera de realizar pruebas adicionales, apoyan la idea de representar la población de mosquitos de forma implícita mediante el parámetro variable en el tiempo  $\beta$ . Finalmente, en los resultados obtenidos para el  $R_0$  en la figura 4.3 se muestran los distintos estados por los que pasa la epidemia. El valor de  $R_0$  llega a sus máximos en los momentos más críticos de la enfermedad, en los que el brote tiende a expandirse y en los que será crucial la intervención de sistemas externos de control. Al final del periodo vemos que la enfermedad tiende por sí misma a la erradicación.

## 4.5. Conclusiones

En este capítulo se ha definido un nuevo modelo para la epidemia de dengue en Cuba mediante ecuaciones diferenciales con retardo. Este modelo presenta la novedad de no incluir de forma explícita datos de las poblaciones de mosquitos, de las que es prácticamente imposible obtener datos fiables, y cuya presencia en los modelos de dengue supone un importante obstáculo a la hora de aplicarlos a casos reales. Para evitar el uso de estas poblaciones se han utilizado retardos y se ha incorporado un parámetro variable en el tiempo,  $\beta$ , que representa la información necesaria sobre esta población y su eficiencia en el contagio. Este parámetro ha sido estimado mediante un método basado en redes neuronales de Hopfield que había sido utilizado previamente para la estimación en modelos epidemiológicos discretos, como se describía en el capítulo 2, pero que no había sido utilizado para sistemas definidos en ecuaciones diferenciales con retardo. Los datos utilizados para realizar la estimación se corresponden con un registro de la epidemia en el brote de La Habana en 2001. Se ha validado la bondad



de la estimación y del propio modelo mediante la comparación de los datos reales y predichos de la epidemia y los resultados han sido satisfactorios. También se ha calculado el valor del número básico de reproducción,  $R_0$ , que ha permitido analizar las distintas fases por las que ha pasado la enfermedad.

De entre las líneas futuras que podrían proponerse en este trabajo, destaca la de llevar a cabo un análisis matemático más profundo del estimador, en el que se estudie de forma rigurosa su convergencia y su comportamiento en presencia de ruido. También sería muy interesante realizar pruebas adicionales con nuevos datos reales de la epidemia, aunque la ausencia de éstos nos lleva a contemplar también la posibilidad de utilizar datos simulados. Finalmente, otra línea de trabajo a proponer sería la de refinar el modelo, ampliándolo de forma que permita obtener una información más directa para realizar medidas de control eficientes. En este sentido, la realización del análisis de sensibilidad del modelo sería un trabajo muy interesante.



## Capítulo 5

# Conclusiones

En esta memoria se ha descrito el modelado de sistemas dinámicos que representan el desarrollo epidemiológico de dos infecciones: VIH/SIDA y dengue. En concreto, se ha mostrado un modelo de la epidemia de VIH/SIDA en Cuba para el que se han estimado algunos de los parámetros más críticos y calculado el número básico de reproducción (capítulo 2). Para esta misma epidemia se ha construido y analizado un modelo novedoso que no incluye explícitamente la población de infectados desconocidos, sino que define la ratio de infección oculta como la razón entre infectados desconocidos y conocidos (capítulo 3). Finalmente, para el caso de la epidemia de dengue ocurrida en La Habana en 2001, se ha definido un modelo con retardos que obvia la necesidad de incluir la población de mosquitos vectores de la enfermedad, procediendo también aquí a realizar la estimación de parámetros y el cómputo del número básico de reproducción (capítulo 4).

En este trabajo se ha contado con dos bases de datos reales anónimas, una para la epidemia de VIH/SIDA y otra para la epidemia de dengue. Para la epidemia de VIH/SIDA, se ha utilizado un registro histórico de 20 años (entre 1986 y 2006), en el que se recogen de forma semanal todos los casos de seropositivos, incluyendo la fecha de detección de la infección, el método utilizado para la misma y, en su caso, las fechas en las que el individuo ha desarrollado SIDA o ha fallecido. Aunque la periodicidad de los datos es semanal, la base de datos ha sido promediada para obtener una periodicidad cuatrimestral, que se

considera suficiente para los objetivos de este trabajo, y que permite filtrar posibles imprecisiones en el registro. La base de datos de la epidemia de dengue utilizada se corresponde con el registro diario de casos del brote ocurrido en La Habana en 2001, debido al virus de serotipo inmunológico DEN-3 [14]. A partir de esa base de datos y de parámetros demográficos y epidemiológicos, tales como la mortalidad o el tiempo de incubación o de recuperación de la enfermedad, se ha construido una base de datos completa en la que se incluyen, además de la población de infectados, la población de individuos susceptibles de contraer la enfermedad y la de individuos recuperados de la misma. Una de las conclusiones que cabe señalar del conjunto del trabajo es la importancia de disponer de datos estadísticos reales sobre la evolución del número de infectados. A diferencia de otras contribuciones que se basan exclusivamente en datos obtenidos de simulaciones, la comparación de los resultados con los datos reales proporciona un apoyo adicional a las conclusiones obtenidas y sirve de validación tanto de los modelos, como de las técnicas de estimación utilizadas.

En el capítulo 2 de la memoria se ha descrito el proceso de identificación de la epidemia de VIH/SIDA en Cuba. Se ha partido de un modelo de la epidemia de VIH/SIDA definido mediante ecuaciones diferenciales ordinarias que había sido descrito en trabajos previos [27]. Este modelo representa una dinámica de poblaciones, cuya característica distintiva es que separa los individuos infectados en diversas poblaciones, en función de si han sido detectados o no, y en este último caso, de cuál ha sido el método de detección. En particular, se define la población de infectados detectados a través del *programa de notificación de parejas*, un sistema de búsqueda activa de personas infectadas de VIH. Con este sistema, cada nuevo individuo registrado como infectado por VIH es instado a declarar sus últimas parejas sexuales, que son a su vez localizadas y sometidas a un test de seropositividad. El análisis de costes y beneficios de este método de detección tiene una gran importancia, ya que es un sistema muy costoso, en el que las autoridades sanitarias han invertido importantes recursos materiales y humanos.

El modelo utilizado se define como un sistema de ecuaciones diferenciales ordinarias que contiene una serie de parámetros que ponderan las transiciones dinámicas entre las distintas poblaciones de la epidemia. De entre ellos destacan los dos parámetros relacionados con la detección de seropositivos, uno correspondiente a la detección mediante el programa de notificación de parejas (denominado  $k_2$ ) y el otro que engloba al resto de métodos de detección (al que se ha llamado  $k_1$ ). Mientras que el resto de parámetros en el modelo son conocidos o pueden al menos estimarse mediante métodos estadísticos, los parámetros de detección son difíciles de medir mediante métodos convencionales. En este trabajo, estos dos parámetros se han considerado variables en el tiempo y han sido calculados utilizando un método de estimación basado en redes neuronales de Hopfield. Este método había sido utilizado con éxito previamente en sistemas mecánicos pero no en sistemas como los epidemiológicos, que no están basados en leyes físicas conocidas sino en el conocimiento de especialistas. Además, tampoco había sido probado con anterioridad en un sistema sin excitación externa. A partir de la estimación de los parámetros también se ha calculado un umbral de detección, que juega un papel similar al del *número básico de reproducción* [19]. Así, este umbral da una idea del estado en el que se encuentra la evolución de la enfermedad, esto es, crecimiento, decrecimiento o endemia. Finalmente, utilizando los resultados obtenidos para la estimación de los parámetros, se han calculado las poblaciones predichas de seropositivos, con el objetivo de compararlas con los datos reales y evaluar la bondad de la estimación y del propio modelo.

En los resultados obtenidos para las estimaciones de los parámetros, podemos observar las diferentes fases por las que ha pasado el valor del parámetro  $k_2$ , que se corresponden con la evolución del programa de notificación de parejas a lo largo del periodo de tiempo estudiado. Mientras que al principio se observa que  $k_2$  tiene un valor relativamente alto, coherente con la apuesta de las autoridades por el programa, entre los años 1992 y 1996 se aprecia un importante descenso, coincidente con el periodo de crisis económica sufrida en el país.

El descenso del valor de los parámetros en este caso no se debe a una menor incidencia de la enfermedad, sino a una menor detección. Finalmente, tras un aumento, en el final del periodo se observa un nuevo descenso del valor del parámetro, relacionado con el fomento de nuevas técnicas de detección, tales como la mayor implicación de médicos de familia. El cálculo del umbral de detección permite observar las mismas fases, siendo el periodo entre 1992 y 1996 el de mayor prevalencia de la enfermedad con valores de umbral significativamente mayores que uno. Para ese periodo, se analiza la relación entre el umbral de detección y el parámetro de detección  $k_2$ , mostrando que disminuir el umbral por debajo de uno (indicando que la epidemia no estaría en crecimiento) requeriría más que duplicar el parámetro. Este incremento de la detección sería inasumible en cuanto a costes económicos y sociales. En resumen, de la discusión de los resultados desde el punto de vista epidemiológico, pueden señalarse las siguientes conclusiones del capítulo 2:

- La detección de infectados mediante el programa de notificación de parejas tiene una influencia significativa sobre la prevalencia de la enfermedad.
- Es preciso complementar el programa de notificación con métodos adicionales de detección, dado que el coste de disminuir significativamente el umbral de detección exclusivamente mediante el programa es inasumible.
- Los resultados que indiquen un descenso en el número de infectados deben tomarse con extrema cautela, ya que pueden deberse a una disminución en la detección, provocando un rebrote en la expansión de la infección que solo se detecte posteriormente.

También dentro del capítulo 2, se utilizaron los resultados de estimación para realizar una comparación entre las poblaciones predichas y reales de la epidemia. Esta comparación mostró un buen ajuste de las curvas, obteniéndose un error cuadrático medio aceptable, lo que sugiere, por un lado, un buen comportamiento del estimador en este caso y, por otro, que el sistema utilizado es un modelo satisfactorio de esta

epidemia. Por otro lado, hay que señalar que se observan fuertes oscilaciones en los valores obtenidos de los parámetros, que se atribuyen a imprecisiones en los datos que afectan la convergencia del estimador. Este problema puede solventarse mediante algoritmos de regularización, característica que se ha incluido variando el número de pasos de integración en la ecuación diferencial del estimador. Los resultados no son definitivos, por lo que este estudio está todavía abierto y la profundización de estas pruebas constituye una prometedora línea futura de trabajo. Pueden señalarse, como resumen del análisis matemático y computacional de este capítulo, las siguientes conclusiones:

- El estimador presentado se ajusta a las características de los sistemas epidemiológicos: sistemas discretos, sin entradas y con parámetros variables.
- Es necesaria la incorporación de mecanismos de regularización.

En el capítulo 3 se propone un modelo de la epidemia de VIH/SIDA, donde la variable es la *ratio de infección oculta*, es decir, el cociente de la población de seropositivos no detectados entre los detectados. La razón de ser de este modelo es que la población de individuos no detectados no aparezca directamente en las ecuaciones. La infección por VIH tiene un periodo de incubación largo y extremadamente variable, por lo que dicha población infectada *oculta* es difícil de estimar de forma fiable. Esta es una limitación fundamental de la mayoría de los modelos publicados, incluido el presentado en el capítulo 2, dado que imposibilita la comparación de los resultados simulados con datos reales. Además, esta población es muy importante desde el punto de vista epidemiológico, ya que la detección temprana de seropositivos permite darles tratamiento e información lo antes posible así como establecer medidas sanitarias para evitar nuevos contagios, lo que supone en sí mismo un eficaz método de control de la enfermedad. Se ha obtenido la solución analítica del modelo de ratio de infección oculta, mostrando que el sistema es asintóticamente estable, lo que debe interpretarse como que, a la larga, la población infectada oculta será proporcional a la detectada. Además, se ha realizado el análisis de es-

tabilidad del sistema, resultando que, con determinadas condiciones, la estabilidad se conservaría incluso aunque el modelo se perturbara moderadamente, resultando una nueva versión del sistema sin solución analítica. Esta *robustez* respecto a perturbaciones es una característica interesante pues sugiere que el comportamiento de la ratio sería cualitativamente igual, aunque la formulación matemática de la detección mediante el programa de notificación de parejas se refinara como se ha sugerido en ocasiones. El valor al que converge la ratio de infección oculta está en torno al 75 %, por encima de la obtenida en trabajos previos que utilizan modelos lineales. También se ha realizado el análisis de sensibilidad del modelo con respecto a los parámetros, obteniendo que el parámetro al que sistema es más sensible es  $k_1$ , el relacionado con los métodos de detección distintos del programa de notificación de parejas. Finalmente, se ha comparado la población real de seropositivos detectados con la población predicha por el modelo, con objeto de evaluar la bondad de este. Se han considerado dos fases distintas de la epidemia, una anterior y otra posterior al año 2000. El objetivo es conseguir un punto inicial coincidente con los datos conocidos y un buen ajuste de la parte final de la curva de la epidemia, así como tener en cuenta que en el 2000 hubo un cambio en la definición de enfermo de SIDA. En ambas fases, especialmente en la segunda, se consigue un buen ajuste del modelo con valores pequeños del error. En resumen, del capítulo 3 se obtienen las siguientes conclusiones:

- Una vez asentada la epidemia, es razonable asumir que el tamaño de la población infectada oculta es, aproximadamente, proporcional a la detectada.
- La ratio de infección oculta es superior a lo que indicarían modelos lineales, lo que sugiere la conveniencia de desarrollar modelos no lineales, que representan más fielmente la evolución de la epidemia.
- El sistema es más sensible al parámetro que mide la detección por métodos adicionales al programa de notificación de parejas, por lo que, al igual que en el capítulo anterior, se manifiesta la



importancia de desarrollar estos métodos, como la implicación de médicos de familia.

En el capítulo 4 se retoma el estimador neuronal basado en redes neuronales de Hopfield que se presentó en el capítulo 2, pero ahora se aplica sobre un nuevo modelo de la epidemia de dengue en Cuba. Este modelo es una dinámica de poblaciones SIR, esto es, sus variables son las poblaciones de individuos susceptibles, infectados y recuperados. La formalización matemática del modelo se realiza mediante ecuaciones diferenciales con retardo. Una de las principales aportaciones de este modelo es que evita el uso explícito de la población del mosquito causante del contagio entre humanos. Esto se consigue mediante la inclusión de retardos que irán desde el tiempo de incubación de la enfermedad en el mosquito hasta su tiempo medio de vida. Por otra parte, se ha definido un parámetro variable que corresponde a la tasa de infección en cada momento, permitiendo modelar indirectamente tanto la población de mosquitos como su eficiencia en el contagio. Esta tasa de infección se estima mediante el estimador basado en redes de Hopfield, que ha sido redefinido para su uso en sistemas con retardos. Los resultados obtenidos para la estimación de la tasa de infección muestran distintas fases de desarrollo de la epidemia. En particular, aparecen valles que no pueden ser explicados por la evolución natural de la enfermedad, sino que deben ser causados por campañas de fumigación exitosas que han interrumpido la expansión de la epidemia. Al igual que se realizó en el capítulo 2, también se ha definido un umbral de detección, que permite observar la fase de evolución en la que se encuentra la epidemia. Los resultados obtenidos para este umbral también apoyan la noción de que la reducción del contagio se debe a las campañas sanitarias, ya que el umbral presenta valores muy altos justo antes del comienzo de la fumigación. Finalmente, para estudiar la bondad de la estimación y del modelo propuesto, se compara la población de infectados de dengue de la base de datos real con la población predicha por el modelo. Para realizar el cálculo de la población predicha, se han integrado las ecuaciones del modelo utilizando los resultados de la estimación de la tasa de infección, observando un

buen ajuste entre las poblaciones predicha y real. En resumen, en el capítulo 4 se destacan las siguientes conclusiones:

- El estimador basado en redes de Hopfield puede ser aplicado a sistemas con retardos, y el buen ajuste obtenido sugiere tanto un comportamiento satisfactorio del estimador como una validación del modelo de dengue propuesto. No obstante, como en el capítulo anterior, se observan fuertes oscilaciones en las estimaciones debidas a la presencia de ruido en los datos, que aconsejan el estudio de las técnicas de regularización.
- Desde el punto de vista epidemiológico, el hallazgo más significativo es la influencia de las medidas sanitarias sobre la evolución de la enfermedad, hasta conseguir su erradicación. Estas medidas de control, que tienen un elevado coste en recursos, han evitado sin embargo que la epidemia se haga endémica, tal como ocurre en otros países de la zona.

El trabajo realizado en esta tesis da pie a considerar varias líneas de desarrollo futuro para la extensión de los resultados aquí obtenidos. En primer lugar, con respecto al modelo de VIH/SIDA presentado en el capítulo 2, sería interesante realizar la estimación de algunos parámetros que en este trabajo se han dado por conocidos. Ya se ha realizado algún trabajo en este sentido [7], en el que se estimó la tasa de incubación de la enfermedad. Otro parámetro que sería interesante estimar con mayor precisión es la tasa de contagio debido a individuos no detectados ya que, por una parte, se presupone que incide significativamente en el desarrollo de la epidemia y, por otro lado, es influenciado directamente por las medidas de control y detección temprana. Asimismo, la realización de un análisis de sensibilidad del modelo para determinar qué parámetros tienen una mayor influencia en la evolución de la epidemia, tal como se ha realizado para el modelo propuesto en el capítulo 3, podría orientar la decisión de qué parámetros incorporar a la estimación, al margen de tener interés por sí mismo. Otra línea interesante que ha quedado fuera del ámbito de este trabajo es el refinamiento del modelo de la epidemia de VIH/SIDA, habiendo aquí

un rango considerable de posibilidades. Podrían considerarse poblaciones separadas en función de circunstancias geográficas, económicas o sociales, o bien podrían proponerse formulaciones matemáticas alternativas del término que modelo el programa de notificación de parejas, incluso como una función genérica desconocida. Hay que señalar que el modelo de la ratio de infección oculta, definido en el capítulo 3, se deriva del de poblaciones presentado en el capítulo 2, por lo que todos los cambios que se realicen en este dan lugar automáticamente a nuevos modelos de infección oculta, cuyo análisis se prevé productivo. Con respecto al modelo del dengue descrito en el capítulo 4, también aquí el estudio de sensibilidad podría ayudar a desarrollar el modelo así como a extraer más información sobre la epidemia y su evolución. Los posibles modelos refinados podrían incorporar explícitamente términos que modelen distintas medidas de control permitiendo estudiar su eficacia relativa. Finalmente, desde un punto de vista matemático y computacional, se propone la realización de un análisis matemático más detallado del estimador neuronal, posiblemente con la incorporación de técnicas de regularización. Se han presentado algunas pruebas preliminares en este sentido [12], y la profundización de este estudio permitiría establecer rigurosamente condiciones para su convergencia y analizar su comportamiento ante la presencia de ruido.



# Capítulo 6

## Conclusions

This thesis has described the process of modelling dynamical systems that represent the epidemic spread of two infections: HIV/AIDS and dengue. In particular, a model of HIV/AIDS in Cuba has been presented, estimating some of the most critical of its parameters and computing the basic reproduction number (Chapter 2). For the same epidemic, a novel model has been built and analyzed, its most prominent feature being that it does not explicitly include the population of unknown infected individuals, but rather a ratio of hidden infection has been defined as the quotient of undetected infected by detected populations (Chapter 3). Finally, for the case of dengue epidemics occurred in Havana in 2001, a model with delays has been defined, dispensing with the need to include the populations of mosquitoes that transmit the virus, proceeding also here to perform parameter estimation and computation of the basic reproduction number (Chapter 4).

This work has relied on two databases of real, anonymized data, one for the epidemic of HIV/AIDS and the other for the dengue epidemic. In the former, we have used a historical record comprising 20 years (from 1986 to 2006), where all infections are weekly registered, including date of infection detection, method used for detection and, where applicable, dates of AIDS diagnosis and decease. Although the periodicity of data is weekly, data have been averaged to obtain three data samples per year, which are enough for the aims of this work, and contributes to filtering eventual data noise. On the other hand, the

dengue epidemic data corresponds to the records of the outbreak occurred in Havana in 2001, due to the virus with immunological serotype 3 [14]. From both that database and demographic and epidemiological parameters, such as mortality, incubation period or recovery period, a sequence has been built comprising not only infected population, but also susceptible and recovered individuals. One of the remarkable conclusions of this work is the importance of having available real statistical data on the evolution of the size of the epidemic. Contrary to other contributions that are solely based upon simulation results, the comparison to real data provides additional support to the obtained conclusions and positively assesses both the models and the estimation techniques.

The process of identification of the epidemic of HIV/AIDS in Cuba is described in Chapter 2. The analysis starts from a model defined by Ordinary Differential Equations that was described in previous work [27]. This model represents a population dynamics, whose distinctive feature is distinguishing the infected individuals according to whether their infection has been detected or not, and in the latter case, according to the method of detection. In particular, the infected population detected through the partner notification program is defined. Such a program implements a policy of active search of HIV infected population, so that newly registered infected individuals are encouraged to disclose their latest sexual partners, who are in turn located and subjected to a seropositive test. A cost-benefit analysis of this program is of utmost importance, since substantial material and human resources have been drained by the program.

The used model is defined as a system of Ordinary Differential Equations that contains several parameters weighting the dynamical transitions between epidemic populations. Among these parameters, those that are related to infection detection stand out, namely the one that models detection by the partner notification program (termed  $k_2$ ) and the parameter that comprises the rest of detection ways (which is called  $k_1$ ). Detection parameters are hardly measurable by conventional methods, whereas the rest of parameters are known, or they can

at least be estimated by statistical methods. In this work, both detection parameters have been considered time-varying, and they have been computed by means of an estimation method based upon Hopfield neural networks. Such a neuronal method had already been successfully applied to mechanical systems, but no experience existed on models that do not stem from physical laws, but rather from the intuition of experts. This is the case of epidemiological systems, which also lack external excitations through inputs. Parameter estimation leads to the computation of a detection threshold, that plays a similar role to that of the *basic reproduction number* [19]. Therefore, such threshold provides an insight on the state of evolution of the epidemic, i.e. increase, decrease, or endemic state. Finally, using the parameters estimates, the predicted infected opulations have been computed, in order to perform a comparison to real data, thus assessing the goodness of both the estimation and the model itself.

The obtained results for parameter estimation show different stages in the value of the parameter  $k_2$ , which correspond to the evolution of the partner notification program over the studied time interval. At the beginning, the estimate of the parameter  $k_2$  is relatively high, which is consistent with the strong support to the program by the health system, whereas between 1992 and 1996 a substantial decrease is discernible, concurring with the economic crisis suffered by the country. The decrease in parameter values, in this case, is not due to a lower disease prevalence, but rather to an impaired detection. Finally, after a slight increase, at the end of the period a new decrease in the parameter estimate is possibly due to the promotion of novel detection ways, such as strengthened implication of general practitioners. The computation of the detection threshold shows similar stages, with the interval between 1992 and 1996 being that of highest prevalence, with threshold values significantly larger than one. For that interval, the relation between the threshold and the detection parameter  $k_2$  is analyzed, showing that decreasing the threshold below one (signalling that the epidemic would be diminishing) would require more than doubling the parameter. This increment would be unaffordable in terms

of economic and social costs. To summarize, from the epidemiological discussion of results, the following conclusions emerge from Chapter 2:

- Infection detection by the partner notification program has a substantial influence on disease prevalence.
- The notification program must be supplemented with additional detection methods, since the cost of a significant reduction of the detection threshold entirely due to the program is unaffordable.
- Those results that suggest a decrease in the size of the infected population must be considered with extreme caution, since they may be due to a weakened detection, causing a reinforced infection outbreak that is only detected afterwards.

Also within the scope of Chapter 2, estimation results were used to compare forecasted and real epidemic populations. The comparison showed a good fit between curves, obtaining acceptable squared mean error, which suggests, on one hand, good estimation performance in this application and, on the other hand, satisfactory modelling of the epidemic system. On another note, strong oscillations often appear in parameter estimates, which are ascribed to data inaccuracies that affect the estimation convergence. This drawback can be alleviated by means of regularization algorithms, a notion that has been incorporated by changing the number of integration steps in the estimator differential equation. Results are far from definitive, though, so this study is still open and deeper analysis is a promising direction for further research. The following conclusions can be pointed out, as a summary of the mathematical and computational analysis of this chapter:

- The presented estimator is well suited to the characteristics of epidemiological systems: discrete systems, without inputs and with time-varying parameters.
- There exists a need for further analysis and implementation of regularization devices.



In Chapter 3, a novel model of HIV/AIDS is proposed by defining a *ratio of hidden infection*, i.e., the quotient of the undetected infected population by the known one. The raison d'être of this model is to avoid the inclusion of the population of infected individuals in the equations. The infection by HIV has a long and extremely variable incubation period, so that the infected *hidden* population is hardly estimated reliably. This is a fundamental shortcoming of most published models, including that presented in Chapter 2, since it hinders comparing simulated results to real data. Besides, managing this population has a strong effect on the epidemic spread, since infected individuals that were detected early could receive treatment and information as soon as possible, and also health measures could be undertaken to avoid new contagions, which is itself an effective way to control the disease. The model of the ratio of hidden infection has been analytically solved, showing that the system is asymptotically stable, which can be interpreted as the proportionality of the undetected infected population to the detected one, in the long term. Besides, the stability analysis shows that, under mild conditions, stability is preserved even under moderate perturbations of the model that would render the system analytically unsolvable. Such a *robustness* with respect to perturbations is an interesting feature, since it suggests that the behaviour of the ratio remains qualitatively similar, even after the mathematical formulation of detection through the partner notification program were refined, as occasionally suggested. The ratio of hidden infection converges to a value that approaches the 75%, beyond that obtained in previous works that used linear models. We have also performed the sensitivity analysis of the model with respect to the parameters, singling out the parameter  $k_1$  as the one that most influences on the system. This somewhat surprising result suggests that the influence of the partner notification program on the hidden infection ratio is limited, compared to alternative detection methods. Finally, the real infected population has been compared to the model prediction, in order to assess the model accuracy. Two epidemic stages have been considered, before and after the year 2000, in order to achieve a good

adjustment both in the initial point and in the final part of the curve, as well as taking into account the change in the definition of AIDS diagnosis that took place in 2000. In both stages, chiefly in the latest, a good model adjustment is observed, with limited error values. To summarize, the following conclusions emerge from Chapter 3:

- Once the epidemic is settled, it is reasonable to assume that the size of the hidden infected population is approximately proportional to the detected infected population.
- The ratio of hidden infection is higher than the estimates of some linear models, which suggests the convenience of paying further attention to sophisticated, non-linear models, that represent more faithfully the epidemic spread.
- The system is more sensitive to the parameter that measures the detection by additional methods excluding the partner notification program, which, as in the previous chapter, puts forth the importance of developing complementary detection methods, such as further implication of general practitioners.

In Chapter 4, the application of the estimator based upon Hopfield neural networks that was presented in Chapter 2 is resumed, but now it is implemented on a novel model of dengue epidemic in Cuba. This model comprises a SIR population dynamics, i.e. its variables represent susceptible, infectious, and recovered (immune) individuals. The mathematical formulation of the model is carried out by means of Delay Differential Equations. A major contribution of this model is the avoidance of populations of mosquitoes explicitly appearing in the model. This is achieved by considering delays that range from the incubation period in the mosquito to its average life span. On the other hand, a variable parameter has been defined to represent the infection rate at each time instant, indirectly modelling both the mosquito populations and their contagion efficacy. Such infection rate is estimated by the estimator based on Hopfield networks, which has been redefined to be applied to systems with delays. The estimates of the infection rate show distinct stages of epidemic spread. In particular,

there appear valleys that cannot be explained by the natural evolution of the disease, but are rather due to successful fumigation campaigns that have disrupted the epidemic spread. As in Chapter 2, a detection threshold has also been defined, providing further insight on the stages that the epidemic evolution traverses. The obtained results for this threshold also support the notion that the reduction in the epidemic spread is due to health measures, since the threshold values are markedly high just before the onset of such campaigns. Finally, in order to assess the goodness of the estimation and the model itself, the infected population from the real database is compared to the population predicted by the model, which results from the numerical integration of the model equations introducing the estimates of the infection rate. A satisfactory fit is observed between real and predicted populations. The following conclusions summarize the results of Chapter 4:

- The estimator based on Hopfield networks can be applied to delay systems, and the obtained good fit suggests both a satisfactory performance of the estimator and a validation of the proposed model for dengue epidemics. However, as in the previous chapter, wide oscillations results from the existence of data noise, thus the analysis of regularization techniques is encouraged.
- From the epidemiological viewpoint, the most significant finding is the influence of health measures on the epidemic evolution, down to disease eradication. These control measures are certainly costly, but they have prevented the onset of endemic dengue, as has settled in other countries of the area.

The work in this thesis raises several directions of future research for further extension of the presented results. Firstly, regarding the model of HIV/AIDS presented in Chapter 2, it would be interesting to undertake the estimation of some parameters that in this work have been assumed as known. Some explorations have already been performed [7], where the incubation rate of the disease was estimated. Another parameter worth being more accurately studied is the conta-

gion rate due to undetected infected individuals since, on one hand, it significantly influences the epidemic spread and, on the other hand, it is directly affected by control and early detection measures. Likewise, a sensitivity analysis of the model, as that performed in Chapter 3, would shed light on which parameters influence most the epidemic evolution, thus guiding the decision on which parameters should be included in the estimation procedure, in addition to being interesting by itself. Another promising direction, which falls beyond the scope of the present work, is the refinement of the model of HIV/AIDS epidemic, existing a considerable range of alternatives. Distinct populations could be considered according to geographic, economic or social circumstances. Otherwise, different mathematical formulations of the term that models the partner notification program could be proposed, even stating a generic unknown function. It is remarkable at this point that the ratio of hidden infection, as defined in Chapter 3, stems from the population model presented in Chapter 2, so every change practiced on the latter automatically yields new models for hidden infection, whose analysis is foreseeably fruitful. Regarding the model of dengue described in Chapter 4, also here the sensitivity analysis could pave the way for further development of the model, as well as provide clearer insight on the epidemic and its evolution. The eventual refined models could explicitly comprise terms that represent diverse control measures, thus providing a way of measuring their relative efficiency. Finally, from a mathematical and computational viewpoint, a deeper theoretical analysis of the estimator is promising, eventually implementing regularization techniques. Preliminary results [12] suggest that this direction of research could lead to rigorous convergence conditions and assess the estimation performance in the presence of data noise.

# Bibliografía

- [1] Abe, S.: 1989, Theories on the Hopfield Neural Networks, *Proc. IEEE International Joint Conference on Neural Networks*, Vol. I, pp. 557–564.
- [2] Arazoza, H., Lounes, R., Pérez, J. and Hoang, T.: 2003, What percentage of the cuban HIV-AIDS epidemic is known?, *Revista Cubana de Medicina Tropical* **55**(1), 30–37.
- [3] Aster, R. C.: 2013, *Parameter estimation and inverse problems*, 2nd ed edn, Academic Press, Waltham, MA.
- [4] Atencia, M.: 2004, *Estabilidad en Sistemas Neuronales Realimentados. Aplicación al Control*, PhD thesis, Departamento de Tecnología Electrónica, Universidad de Málaga.
- [5] Atencia, M., García-Garaluz, E., de Arazoza, H. and Joya, G.: 2013, Estimation of parameters based on artificial neural networks and threshold of HIV/AIDS epidemic system in Cuba, *Mathematical and Computer Modelling* **57**(11-12), 2971–2983.
- [6] Atencia, M., García-Garaluz, E. and Joya, G.: 2013, The ratio of hidden HIV infection in Cuba, *Mathematical Biosciences and Engineering* **10**(4), 959–977.
- [7] Atencia, M., Joya, G., García-Garaluz, E., de Arazoza, H. and Sandoval, F.: 2007, Estimation of the Rate of Detection of Infected Individuals in an Epidemiological Model, *in* F. Sandoval, A. Prieto, J. Cabestany and M. Graña (eds), *Computatio-*

- nal and Ambient Intelligence, Lecture Notes In Computer Science (IWANN 2007)*, Vol. 4507, Springer, pp. 948–955.
- [8] Atencia, M., Joya, G. and Sandoval, F.: 2003, Modelling the HIV-AIDS Cuban Epidemics with Hopfield Neural Networks, *in* J. Mira and J. R. Álvarez (eds), *Artificial Neural Nets. Problem Solving Methods. Lecture Notes in Computer Science (IWANN 2003)*, Vol. 2687, Springer, pp. 449–456.
- [9] Atencia, M., Joya, G. and Sandoval, F.: 2004, Parametric Identification of Robotic Systems with Stable Time-varying Hopfield Networks, *Neural Computing & Applications* **13**(4), 270–280.
- [10] Atencia, M., Joya, G. and Sandoval, F.: 2005a, Dynamical Analysis of Continuous Higher Order Hopfield Networks for Combinatorial Optimization, *Neural Computation* **17**(8), 1802–1819.
- [11] Atencia, M., Joya, G. and Sandoval, F.: 2005b, Hopfield Neural Networks for Parametric Identification of Dynamical Systems, *Neural Processing Letters* **21**(2), 143–152.
- [12] Atencia, M., Joya, G. and Sandoval, F.: 2013, Identification of noisy dynamical systems with parameter estimation based on Hopfield neural networks, *Neurocomputing* **121**, 14–24.
- [13] Bailey, N. T.: 1975, *The mathematical theory of infectious diseases and its applications*, Ch. Griffin and Company LTD.
- [14] Barrios, J., Piétrus, A., Joya, G., Marrero, A. and Arazoza, H.: 2013, A differential inclusion approach for modeling and analysis of dynamical systems under uncertainty: Application to dengue disease transmission, *Soft Computing* **17**(2), 239–253.
- [15] Bernoulli, D.: 1766, Essai d’une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l’inoculation pour la prévenir, *Mémoires de mathématique et de physique* .
- [16] Coutinho, F., Burattini, M., Lopez, L. and Massad, E.: 2006, Threshold conditions for a non-autonomous epidemic system descri-

- bing the population dynamics of dengue, *Bulletin of Mathematical Biology* **68**(6), 2263–2282.
- [17] Dickinson, R. P. and Gelinás, R. J.: 1976, Sensitivity analysis of ordinary differential equation systems—a direct method, *Journal of Computational Physics* **21**(2), 123–143.
- [18] Diekmann, O., Heesterbeek, J. and Metz, J.: 1990, On the definition and the computation of the basic reproduction ratio  $r_0$  in models for infectious diseases in heterogeneous populations, *Journal of Mathematical Biology* **28**(4).
- [19] Diekmann, O. and Heesterbeek, J.A.P.: 2000, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, Wiley.
- [20] Engl, H. W., Hanke, M. and Neubauer, A.: 2000, *Regularization of inverse problems*, Kluwer Academic Publishers.
- [21] Esteva, L. and Vargas, C.: 1999, A model for dengue disease with variable human population, *Journal of Mathematical Biology* **38**(3), 220–240.
- [22] García-Garaluz, E., Atencia, M., García-Lagos, F., Joya, G. and Sandoval, F.: 2009, System identification of dengue fever epidemics in cuba, in J. Cabestany, F. Sandoval, A. Prieto and J. M. Corchado (eds), *Bio-Inspired Systems: Computational and Ambient Intelligence*, Vol. 5517, Springer Berlin Heidelberg, pp. 901–908.
- [23] García-Garaluz, E., Atencia, M., Joya, G., García-Lagos, F. and Sandoval, F.: 2011, Hopfield networks for identification of delay differential equations with an application to dengue fever epidemics in cuba, *Neurocomputing* **74**(16), 2691–2697.
- [24] Hagiwara, K.: 2002, Regularization learning, early stopping and biased estimator, *Neurocomputing* **48**(1-4), 937–955.

- [25] Hirsch, M. W. and Smale, S.: 1974, *Differential Equations, Dynamical Systems, and Linear Algebra*, Academic Press.
- [26] Hopfield, J.: 1982, Neural Networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558.
- [27] Hsieh, Y.-H., Arazoza, H., Lounes, R. and Joanes, J.: 2005, A Class of Methods for HIV Contact Tracing in Cuba: Implications for Intervention and Treatment, in W.-Y. Tan and W. Hulin (eds), *Deterministic and Stochastic Models of AIDS Epidemics and HIV Infections with Intervention*, World Scientific, pp. 77–92.
- [28] Hsieh, Y.-H., Wang, K., Wang, H.-C., Arazoza, H. D., Lounes, R., Twu, S.-J. and Hsu, H.-M.: 2005, Ascertaining HIV Underreporting in Low Prevalence Countries using the Approximate Ratio of Underreporting, *Journal of Biological Systems* **13**(4), 441 – 454.
- [29] Ibrahim, F., Taib, M. N., Abas, W. A. B. W., Guan, C. C. and Sulaiman, S.: 2005, A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN), *Computer methods and programs in biomedicine* **79**(3), 273–281.
- [30] Joya, G., Atencia, M. and Sandoval, F.: 2002, Hopfield Neural Networks for Optimization: Study of the Different Dynamics, *Neurocomputing* **43**(1-4), 219–237.
- [31] Khalil, H. K.: 2002, *Nonlinear Systems*, Prentice Hall.
- [32] Ljung, L.: 1999, *System Identification. Theory for the User*, Prentice Hall.
- [33] Murray, J.: 2002, *Mathematical biology*, 3rd ed. edn, Springer, New York.
- [34] Naresh, R., Tripathi, A. and Sharma, D.: 2011, A nonlinear HIV/AIDS model with contact tracing, *Applied Mathematics and Computation* **217**(23), 9575–9591.



- [35] Rapatski, B., Klepac, P., Dueck, S., Liu, M. and Weiss, L. I.: 2006, Mathematical epidemiology of HIV/AIDS in cuba during the period 1986-2000, *Mathematical Biosciences and Engineering* **3**(3), 545–556.
- [36] Schittkowski, K.: 2002, *Numerical Data Fitting in Dynamical Systems*, Kluwer Academic Publishers.
- [37] Slotine, J.-J. and Li, W.: 1991, *Applied Nonlinear Control*, Prentice Hall.
- [38] Tan, W.-Y. and Hulin, W. (eds): 2005, *Deterministic and stochastic models of AIDS epidemics and HIV infections with intervention*, World Scientific, Singapore ;;Hackensack NJ.
- [39] Tank, D. and Hopfield, J.: 1985, 'Neural' Computation of Decisions in Optimization Problems, *Biological Cybernetics* **52**, 141–152.
- [40] Unbehauen, H.: 1996, Some New Trends in Identification and Modeling of Nonlinear Dynamical Systems, *Applied Mathematics and Computation* **78**, 279–297.
- [41] Vidyasagar, M.: 2002, *Minimum-seeking properties of analog neural networks with multilinear objective functions*, Society for Industrial and Applied Mathematics.
- [42] World Health Organization: 2014, Fact sheet 117: Dengue and severe dengue.  
**URL:** <http://www.who.int/mediacentre/factsheets/fs117/en/>



# Apéndice A

## Publicaciones que compendian la tesis

- A.1. Estimation of parameters based on artificial neural networks and threshold of HIV/AIDS epidemic system in Cuba





Contents lists available at SciVerse ScienceDirect

# Mathematical and Computer Modelling

journal homepage: [www.elsevier.com/locate/mcm](http://www.elsevier.com/locate/mcm)



## Estimation of parameters based on artificial neural networks and threshold of HIV/AIDS epidemic system in Cuba<sup>☆</sup>



M. Atencia<sup>b,\*</sup>, E. García-Garaluz<sup>a</sup>, H. de Arazoza<sup>c</sup>, G. Joya<sup>a</sup>

<sup>a</sup> Departamento de Tecnología Electrónica, E.T.S.I.Telecomunicación, Universidad de Málaga, Campus de Teatinos, 29071 Málaga, Spain

<sup>b</sup> Departamento de Matemática Aplicada, E.Politécnica Superior, Universidad de Málaga, Campus de Teatinos, 29071 Málaga, Spain

<sup>c</sup> Facultad de Matemática y Computación, Universidad de La Habana, San Lázaro y L, Vedado, La Habana, Cuba

### ARTICLE INFO

#### Article history:

Received 26 January 2010

Received in revised form 3 February 2013

Accepted 8 March 2013

#### Keywords:

Epidemic model  
Hopfield neural networks  
Dynamical systems  
Systems identification

### ABSTRACT

In this paper, a method of parameter estimation, based upon Hopfield neural networks, is applied to the identification of a model of the HIV/AIDS epidemic in Cuba. This estimation technique presents a number of features that make it especially suitable for the identification of such epidemic models. In particular, its remarkable ability to estimate time variant parameters allows us to outline the evolution of parameters, thus assessing the efficiency of health policies. A complementary aim is the validation of the model, which stems from basic concepts and reasoning in disease modelling rather than from specific physical laws, thus it is not supported by rigorous mathematical results. Indeed, the experimental results show a significant adjustment of simulated data to actual ones, which provides considerable support to model validity. A sequence of real data, corresponding to the HIV and AIDS affected populations in Cuba from 1986 to 2004, has been compared to both simulations of the epidemic model and used to test the estimation method. A detection threshold, which plays a similar role to the basic reproduction number, has been calculated to complement the information about the current situation of the epidemic and the possible evolution perspectives. The obtained results are plausible and coherent according to health experts.

© 2013 Elsevier Ltd. All rights reserved.

### 1. Introduction

The main objective of this paper is the estimation of the parameters appearing in a HIV/AIDS epidemic model in the Cuban case. In particular, the parameters related to detection rates, which are time variant, are estimated by a modified version of Hopfield Neural Networks. The knowledge of the numerical value of these parameters allows for computing an estimation of a threshold that can be interpreted in the same way as the basic reproduction number, hence it provides important information about the eventual future evolution of the epidemic spread. Besides, the plausible simulation results provide support to the model validity.

Reality modelling is a fundamental task in the engineering framework. In systems with no previous information, system identification, defined as the determination of its internal properties from measurable external outputs [1], is a feasible way of building a system model. Studies and mathematical modelling of infectious diseases have historically been carried out

This work has been partially supported by the Agencia Española de Cooperación Internacional para el Desarrollo (AECID) from Spain, project no. A2/038418/11. Thanks are due to Dr. Inma P. Cabrera for useful stylistic suggestions. The thorough reading and useful suggestions of the reviewers are gratefully acknowledged.

Corresponding author. Tel.: +34 951 952 423; fax: +34 95 213 27 66.

E-mail addresses: [matencia@ctima.uma.es](mailto:matencia@ctima.uma.es), [matencia@uma.es](mailto:matencia@uma.es) (M. Atencia).

0895-7177/\$ – see front matter © 2013 Elsevier Ltd. All rights reserved.

<http://dx.doi.org/10.1016/j.mcm.2013.03.007>

## A.2. The ratio of hidden HIV infection in Cuba



## THE RATIO OF HIDDEN HIV INFECTION IN CUBA

MIGUEL ATENCIA

Dept. of Applied Mathematics, University of Málaga  
29071 Málaga, Spain

ESTHER GARCÍA-GARALUZ AND GONZALO JOYA

Dept. of Electronics Technology, University of Málaga  
29071 Málaga, Spain

(Communicated by Yang Kuang)

**ABSTRACT.** In this work we propose the definition of the ratio of hidden infection of HIV/AIDS epidemics, as the division of the unknown infected population by the known one. The merit of the definition lies in allowing for an indirect estimation of the whole of the infected population. A dynamical model for the ratio is derived from a previous HIV/AIDS model, which was proposed for the Cuban case, where active search for infected individuals is carried out through a contact tracing program. The stability analysis proves that the model for the ratio possesses a single positive equilibrium, which turns out to be globally asymptotically stable. The sensitivity analysis provides an insight into the relative performance of various methods for detection of infected individuals. An exponential regression has been performed to fit the known infected population, owing to actual epidemiological data of HIV/AIDS epidemics in Cuba. The goodness of the obtained fit provides additional support to the proposed model.

**1. Introduction.** The main objective of this work is to study the proportion of infected individuals who are undetected over the detected infected population in a contagious disease, as an indirect methodology to provide an insight on the asymptomatic population. These individuals are often unaware of their own infection, thus they do not take prophylactic measures to avoid contagion to susceptible individuals. Information about undetected infected people is especially interesting for the HIV/AIDS epidemics since this asymptomatic stage can extend over time, avoiding early assistance and allowing for a greater spread of the disease. Indeed, the incubation period of AIDS is rather variable and, in contrast to other viral infections, the seropositive individual never ceases to be contagious, which further complicates the control of the epidemic spread. Consequently, a major topic for health systems is developing new strategies for early detection, that is to say minimizing the size of the undetected infected population. These strategies are expensive both in human and economic resources, thus the study of their efficiency is a key point.

Mathematical modelling of infectious diseases as dynamical systems has received considerable scientific interest along history (see e.g. [6, 2, 22]). Epidemiological models are often constructed upon systems of Ordinary Differential Equations

---

2010 *Mathematics Subject Classification.* Primary: 34D20, 93E12; Secondary: 92D30.

*Key words and phrases.* HIV/AIDS epidemiology, population dynamics, dynamical systems, Lyapunov stability, system identification.



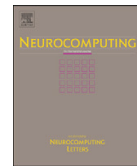
### A.3. Hopfield networks for identification of delay differential equations with an application to dengue fever epidemics in Cuba





Contents lists available at ScienceDirect

## Neurocomputing

journal homepage: [www.elsevier.com/locate/neucom](http://www.elsevier.com/locate/neucom)Hopfield networks for identification of delay differential equations with an application to dengue fever epidemics in Cuba<sup>☆</sup>Esther García-Garaluz<sup>a</sup>, Miguel Atencia<sup>b,\*</sup>, Gonzalo Joya<sup>a</sup>, Francisco García-Lagos<sup>a</sup>, Francisco Sandoval<sup>a</sup><sup>a</sup> Departamento de Tecnología Electrónica, Universidad de Málaga (Spain) Campus de Teatinos, 29071 Málaga, Spain<sup>b</sup> Departamento de Matemática Aplicada, Universidad de Málaga (Spain) Campus de Teatinos, 29071 Málaga, Spain

## ARTICLE INFO

Available online 12 May 2011

## Keywords:

Parameter estimation  
 Dynamical system  
 Hopfield neural networks  
 Delay differential equation  
 Dengue fever epidemic

## ABSTRACT

This work is aimed at proposing an algorithm, based upon Hopfield networks, for estimating the parameters of delay differential equations. This neural estimator has been successfully applied to models described by Ordinary Differential Equations, whereas its application to systems with delays is a novel contribution. As a case in point, we present a model of dengue fever for the Cuban case, which is defined by a delay differential system. This epidemiological model is built upon the scheme of an SIR (susceptible, infected, recovered) population system, where both delays and time-varying parameters have been included. The latter are thus estimated by the proposed neural algorithm. Additionally, we obtain an expression of the Basic Reproduction Number for our model. Experimental results show the ability of the estimator to deal with systems with delays, providing plausible parameter estimations, which lead to predictions that are coherent with actual epidemiological data. Besides, when the Basic Reproduction Number is computed from the estimated parameter values, results suggest an evolution of the epidemic that is consistent with the observed infection. Hence the estimation could help health authorities to both predict the future trend of the epidemic and assess the efficiency of control measures.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

System identification can be defined as the determination of the internal properties of a system from measurable outputs [1]. In models that contain parameters, fitting the values of the model parameters is referred to as parametric identification or parameter estimation [2]. In this contribution, we describe a method for parameter estimation, based on Hopfield neural networks. The proposed algorithm presents two advantages over conventional techniques: it allows for a time-variant estimation of the parameters, and it can be formulated to estimate parameters of systems formulated by delay differential equations (DDE).

The application of mathematics to epidemiological modelling is a task that has been performed for a long time [3,4]. Building models of infectious diseases, validating them and estimating their parameters, allows, on one hand, to predict their evolution, and on the other hand, to develop sanitary policies in order to fight the diseases. Dengue fever is an infectious disease which has reappeared in subtropical

regions of the American and African continent in the last decades. Therefore, modelling and identification of the dengue fever system have become an interesting study to be performed [5]. It has also been investigated the phenomenon of dengue overwintering [6], by providing interesting hypotheses on mosquito life cycle. These efforts, which have helped to construct a valuable theoretical framework for dengue analysis, face however a fundamental shortcoming when it comes to comparing the model to actual statistical data of the infection: the inclusion in the model of mosquito populations. There is no current or foreseeable way to measure the number of mosquitos that live in an area, let alone determine how many of them are infected by the virus. Consequently, the first objective of the paper is posing a model of dengue fever that dispenses with the explicit inclusion of mosquito populations. Conventional models of dengue represent the complete process of human–mosquito–human virus transmission by defining separate populations of humans and mosquitos. Thus, there is an explicit passage of mosquitos to an infected state, which in turn leads to a passage of healthy humans to infected individuals. In contrast, the presented model [7] only includes human populations, whereas the effectiveness of disease transmission by mosquitos is accounted for by means of a mathematical expression, which includes a parameter that somehow measures the whole cycle from an infected human to a susceptible individual being infected. Due to the incubation period of the virus within the mosquito, the new infections are due to mosquitos that became infected in the past, thus the included parameter must be related to past values of

This work has been partially supported by the Spanish Ministerio de Ciencia e Innovación (Project no. TIN2008-04985), by the Junta de Andalucía (Project no. P08-TIC-04026), and Agencia Española de Cooperación Internacional para el Desarrollo (Projects no. D/017218/08, D/023835/09 and D/030223/10). The useful suggestions of the anonymous reviewers are gratefully acknowledged.

Corresponding author. Fax: +34 95 213 27 66.  
 E-mail address: [matencia@ctima.uma.es](mailto:matencia@ctima.uma.es) (M. Atencia).

0925-2312/\$ - see front matter © 2011 Elsevier B.V. All rights reserved.  
 doi:10.1016/j.neucom.2011.03.022