



UNIVERSIDAD DE LA HABANA
FACULTAD DE MATEMÁTICAS Y COMPUTACIÓN

**MODELACIÓN DE DATOS DE BROTES DE DENGUE,
ESTIMACIÓN Y PREDICCIÓN DE PARÁMETROS
EPIDEMIOLÓGICOS**

**TESIS EN OPCIÓN AL GRADO CIENTÍFICO
DOCTOR EN CIENCIAS MATEMÁTICAS**

AUTOR: CARLOS RAFAEL SEBRANGO RODRÍGUEZ

LA HABANA, CUBA.

2018



UNIVERSIDAD DE LA HABANA
FACULTAD DE MATEMÁTICAS Y COMPUTACIÓN

**MODELACIÓN DE DATOS DE BROTES DE DENGUE,
ESTIMACIÓN Y PREDICCIÓN DE PARÁMETROS
EPIDEMIOLÓGICOS**

**TESIS EN OPCIÓN AL GRADO CIENTÍFICO
DOCTOR EN CIENCIAS MATEMÁTICAS**

AUTOR: CARLOS RAFAEL SEBRANGO RODRÍGUEZ

TUTORES: Dr. C. LIZET SÁNCHEZ VALDÉS
Dr. C. ZIV SHKEDY
Dr. C. VIVIAN SISTACHS VEGA

LA HABANA, CUBA.

2018

Agradecimientos

A mis tutores Lizet Sánchez, Ziv Shkedy y Vivian Sistach por sus enseñanzas, estímulo incansable, consejos y guía en la realización de esta investigación.

A mis compañeros de trabajo de la Universidad de Sancti Spíritus "José Martí Pérez" en especial a los profesores del Departamento de Matemática-Física (Catalina, Ricardo, Carbonell, Tania, Yunia, Mayra, . . . , etc.) y del CEEPI por su apoyo y confianza durante estos años.

A los colegas del Instituto de Medicina Tropical IPK "Pedro Kourí", en especial a Marta Castro, Denis, Lupe, . . . , etc. por su apoyo, sugerencias y explicaciones epidemiológicas.

A Ivonne Barlia y Ewoud De Troyer por su ayuda para comprender la programación en R.

A Emanuele, Yovanna, Fortunato, Donato, Consuelo, Martin, Terylee, Ariel, Sammy, Wiebke, Trishanta, . . . por hacer mis estancias en Bélgica mucho más placentera, especialmente cuando extrañaba mucho a mi familia.

A mis amigos y compañeros de estudio, aunque estén lejos.

A todos los miembros de mi familia, en particular, a mis padres por inculcarme el amor a las matemáticas, por la educación, por su infinito amor; a mi hermano, a Alejandrino, a Mileidi, a Victor, a mis suegros y otros por su apoyo emocional y aliento.

A mi esposa, por su amor, su paciencia, por cuidar nuestros hijos en mis visitas a Bélgica y en mis horas de estudio, por su constante apoyo y aliento. Sin ella, sin dudas, habría sido mucho más difícil terminar el doctorado.

A todos, **Muchas Gracias.**

A la memoria de mi padre, JULIO RAFAEL SEBRANGO ALONSO,
A mi madre, OLANIA RODRÍGUEZ PEÑA,
A mi esposa ROSABELL PÉREZ GUTIERREZ
A mis hijos CARLOS ERNESTO, DIANA ELENA y DIEGO ALEJANDRO,
A mi hermano EDUARDO.

”Una cuestión que concierne de tan cerca el bien de la humanidad, no debe ser otra cosa que todo el conocimiento de las causas que nos puede proporcionar un poco de análisis y de cálculo”.

Daniel Bernoulli, 1760.

SÍNTESIS

A pesar de la popularidad de los modelos matemáticos de transmisión del dengue, estos se ven obstaculizado por el hecho de que no siempre hay demasiada información disponible. Una alternativa en esta situación es hacer uso de los modelos fenomenológicos. La investigación se centra en el estudio de tales modelos para la estimación y predicción de parámetros epidemiológicos claves. En la presente tesis se proponen nuevas alternativas metodológicas para la modelación de datos de brotes de dengue. Para brotes de una onda se propone un procedimiento que utiliza el promedio de modelos y se proponen diferentes enfoques metodológicos para tener en cuenta la incertidumbre de modelos en la comparación de brotes. Además, se aplican los modelos no lineales con efectos mixtos para el estudio de la heterogeneidad de la transmisión entre diferentes áreas de una misma región. Para la modelación de brotes multiondas se propone una nueva alternativa metodológica utilizando la verosimilitud perfil. Finalmente, se hace una introducción y se visualizan las funciones del paquete de R creado, *DengueRT*, que permitió implementar computacionalmente la metodología propuesta para brotes de una onda. Todas estas alternativas fueron aplicadas a datos de brotes de dengue ocurridos en Cuba.

Índice

Introducción	1
Objetivos	5
Objetivo general	5
Objetivos específicos	5
Novedad científica	5
Estructura de la tesis	6
1 El dengue y la modelación de brotes de dengue	8
1.1 Introducción	8
1.2 Dengue	9
1.3 Modelos Matemáticos para brotes de dengue	10
1.3.1 Modelos de transmisión	10
1.3.2 Modelos fenomenológicos	13
1.4 Descripción de los datos en la modelación	18
1.5 Conclusiones parciales	19
2 Incertidumbre de modelos de brotes de una onda	20
2.1 Introducción	20
2.2 Modelos no lineales y métodos de estimación	21
2.2.1 Métodos de estimación	22
2.3 Incertidumbre, selección y promedio de modelos	28
2.3.1 Cuantificación de la incertidumbre de los parámetros	31

2.3.2	Propuesta de procedimiento	32
2.4	Aplicación de la metodología	33
2.4.1	Estimación usando métodos del promedio de modelos	33
2.4.2	Predicción en tiempo real	35
2.5	Conclusiones parciales	37
3	Modelación de brotes de una onda en diferentes regiones	39
3.1	Introducción	39
3.2	Incertidumbre de modelos en la comparación de dos brotes	40
3.2.1	Enfoque Frecuentista	40
3.2.2	Enfoque Bayesiano	42
3.2.3	Aplicación: Comparación brotes dengue, La Lisa 2006 y 2007	47
3.3	Modelación de datos de varias áreas en una misma región	50
3.3.1	Metodología clásica: Modelo para cada área específica	51
3.3.2	Metodología propuesta: Modelo no lineal con efectos mixtos	52
3.3.3	Aplicación de las metodologías	58
3.4	Conclusiones parciales	63
4	Modelación de brotes multiondas	65
4.1	Introducción	65
4.2	Modelación utilizando verosimilitud perfil	66
4.2.1	Modelación de brotes de dos ondas	66
4.2.2	Modelación de brotes de tres ondas epidémicas ó más	69
4.3	Propuesta de procedimiento	70
4.4	Ejemplos de aplicación	71
4.4.1	Utilizando verosimilitud perfil	72
4.4.2	Evaluación del impacto de una intervención	73
4.5	Conclusiones parciales	77
5	Desarrollo de Software: Paquete de R, <i>DengueRT</i>	79
5.1	Introducción al paquete <i>DengueRT</i>	80

5.1.1	Implementación de las funciones y datos	80
5.1.2	Creación del paquete <i>DengueRT</i>	82
5.1.3	Uso de la plataforma <i>R-forge</i>	86
5.2	Visualización de las funciones del paquete <i>DengueRT</i>	89
5.2.1	Estimación de parámetros. Ejemplo de aplicación en dengue	89
5.2.2	Predicciones en tiempo real. Ejemplo de aplicación en dengue	92
5.3	Conclusiones parciales	97
	Conclusiones	99
	Recomendaciones	100
	Referencias	101
	Anexos	111

Introducción

En los últimos años se ha incrementado el uso de los modelos matemáticos para la modelación de enfermedades infecciosas [16,59]. Estos modelos han permitido describir el comportamiento de brotes, facilitar la estimación de parámetros epidemiológicos claves, evaluar el impacto de las intervenciones de control, hacer predicciones en tiempo-real y ayudan a entender y describir la dinámica de la transmisión. Fundamental en este incremento ha sido la integración con métodos estadísticos rigurosos.

Dentro de las enfermedades infecciosas beneficiadas por la modelación matemática está el dengue. El dengue es la infección viral más importante transmitida por artrópodos que afecta al humano [32] y ha emergido como un importante problema de salud en las regiones tropicales y subtropicales. Actualmente es endémica, con brotes epidémicos en América, el sudeste de Asia y el Pacífico [31]. Esta enfermedad es causada por cuatro serotipos virales y es transmitida por la picada de un mosquito *Aedes aegypti* infectado. Los patrones de reproducción de la enfermedad en las poblaciones humanas están influenciados por una gran variedad de factores entre los que se encuentran el clima, factores ambientales y de urbanización, la densidad poblacional, problemas con el sistema de recolección de basura y el suministro de agua que favorecen la proliferación de sitios de cría del vector. Las acciones de control, al no existir vacuna ni tratamiento específico para el dengue, se basan fundamentalmente en el control del mosquito y en la vigilancia epidemiológica activa de la enfermedad y sus factores de riesgo [32].

Cuba cuenta con un Sistema Nacional de Salud bien estructurado y presenta un programa de control de *Aedes aegypti* que monitorea y controla sistemáticamente las poblaciones de este mosquito, manteniéndolo en niveles comparativamente bajos con respecto al mundo. Sin

embargo, a pesar de esto, sucesivos brotes han afectado al país desde el 2005 en diferentes localidades debido principalmente a la presencia del vector y la afluencia constante de viajeros provenientes de áreas endémicas de dengue. La ocurrencia de la enfermedad en forma de brotes, con la circulación fundamentalmente de un solo serotipo, diferente en cada momento, así como la alta susceptibilidad de la población, hace que nuestras condiciones epidemiológicas sean únicas [32]. El reporte semanal de casos que se registra a los niveles de municipio y provincia constituyen la principal fuente de información para la modelación.

Los modelos matemáticos de transmisión ponen diferentes énfasis en cada uno de estos aspectos de la epidemiología del dengue, reflejando sus intereses y preguntas específicas y están diseñados para abordar preguntas específicas que pueden ser clasificadas en cinco categorías: (1) cuantificar la transmisión en diferentes entornos; (2) comprender los patrones temporales de la enfermedad; (3) predecir el patrón espacial de transmisión; (4) comprender la estructura serotipo / cepa; y (5) estimar el impacto potencial de los métodos de control [31].

A pesar de la popularidad en el uso de los modelos de transmisión del dengue, estos se ven obstaculizados por la notable incertidumbre sobre la epidemiología de la enfermedad, la dificultad de estar en un contexto de datos limitados y la ausencia de información confiable sobre los mecanismos de transmisión [14]. Una alternativa en esta situación es hacer uso de los modelos fenomenológicos. Los modelos fenomenológicos (por ejemplo, Hsieh y cols. [38]) prescindan de los supuestos de los modelos de transmisión. En su lugar, ellos hacen supuestos generales acerca de la forma de la curva de incidencia. En general, utilizan pocos parámetros, hacen suposiciones más sencillas, y probablemente dan estimaciones más robustas cuando son aplicados en contextos con datos muy limitados [54]. Nuestra investigación se centra en el estudio de tales modelos para la estimación y predicción de parámetros epidemiológicos claves.

Entre los modelos fenomenológicos más utilizado para la modelación de datos de enfermedades infecciosas se encuentra el modelo de Richards [35–43,70,85]. El modelo de Richards considera sólo el tamaño de la población infecciosa acumulada con saturación en el crecimiento cuando el brote avanza. La premisa básica del modelo de Richards es que la curva de incidencia está compuesta de un único pico de alta incidencia, lo que resulta en una curva epidémica en forma

de S [35, 36, 42, 43]. El modelo de Richards se ha utilizado con frecuencia para modelar los casos notificados acumulados en los brotes de dengue. En particular, Hsieh y cols. [35, 42] propuso utilizar el modelo de Richards para estimar el tamaño final, el punto de retorno o acmé de la epidemia y el número reproductivo básico R_0 durante un brote de una onda. Sin embargo, otros modelos no lineales de crecimiento (Logístico de 3 parámetro [27, 35, 83], logístico de 5 parámetro [50], Sigmoidal Emax [55], Gompertz [27, 83] y Weibull [77]) pudieran ajustar bien estos datos epidémicos. Hasta ahora, ninguno de ellos ha sido empleado para estimar parámetros epidemiológicos y hacer predicciones en tiempo real en brotes de dengue.

La selección de modelos es un tema central en la modelación estadística. Sin embargo, a menudo nos enfrentamos con el problema de la existencia de varios modelos que pueden ser ajustados a un mismo conjunto de datos. Típicamente la selección del modelo de mejor ajuste dentro de todos los modelos posibles ignora la incertidumbre de la selección de modelos en la estimación e inferencia. Por estas razón, varios autores, entre los cuales están Burnham & Anderson [10], Posada [65], Moon [57], Claeskens & Hjort [18] y Lin [51] abogaron por el uso de técnicas de promedio de modelos (*model averaging*) para realizar inferencias multi-modelos en los parámetros estimados. El promedio de modelos es un método que tiene en cuenta todos los modelos ajustados para la estimación de los parámetros de interés. Está basado en el promedio ponderado de los parámetros estimados obtenido de los diferentes modelos, dándole mayor peso a los modelos que mejor ajustan los datos [26, 45, 58]. Esta pudiera ser una alternativa para la estimación y predicción en tiempo real de parámetros epidemiológicos en brotes de dengue y también para la comparación de brotes a través de estos parámetros, aspecto poco abordado en la literatura.

En ocasiones el número de casos acumulados reportados se colecciona para todas las áreas urbanas en una región en particular. Muchas veces resulta de interés estimar parámetros epidemiológicos claves o estudiar la heterogeneidad o variabilidad regional de la transmisión de la enfermedad a través del número reproductivo básico R_0 . La modelación, para estimar parámetros epidemiológicos en estas situaciones, se lleva a cabo usualmente ajustando el modelo de Richards individual para cada área específica por separado. Una limitación de este

abordaje es que no se tiene en cuenta la estructura urbana y se ignora la variabilidad entre las áreas. Una forma eficaz de abordar la modelación de datos de brotes cuando interesa tener en cuenta la estructura urbana de la población en una región, es decir, la transmisión entre áreas en una región determinada, o estimar el comportamiento promedio de una área, y la variabilidad entre y dentro de las áreas, sería utilizando un modelo no lineal con efectos mixtos [6, 23]. Sin embargo, no han sido utilizados hasta el momento para la modelación de este tipo de datos en brotes de dengue.

En la modelación de datos de brotes de dengue cuando hay más de una onda de infecciones, se ha propuesto el empleo de una variación del modelo de Richards: El modelo de Richards multifases. Recientemente, Hsieh [35, 38–40] demostró su utilización para modelar brotes multiondas de SARS [40], dengue [38, 39], Zika [37], pH1N1 [43] y Ébola [36]. Este procedimiento consiste en el ajuste del modelo de Richards a los casos acumulados en unidades de tiempo sucesivas, si la estimación del parámetro que indica el pico de la epidemia comienza a divergir desde un valor convergente previo, esto indicaría el inicio de otra onda epidémica, cuya estimación se obtiene determinando el mínimo local de la curva de incidencia. Luego, se puede ajustar el modelo de Richards a la curva de casos acumulados nuevamente, pero comenzando desde la unidad de tiempo después del inicio de la segunda fase. Este proceso se repite hasta que termine el brote. Una limitación en este procedimiento es que la estimación de los parámetros en cada una de las ondas no se hace simultáneamente, sino que considera, de forma separada los segmentos sucesivos de la curva epidémica para la estimación de los parámetros del modelo en cada una de las ondas. Lo que motiva a la búsqueda de alternativas metodológicas que consideren el uso de modelos no lineales que permitan la estimación de los parámetros epidemiológicos teniendo en cuenta los segmentos de la curva epidémica simultáneamente.

Teniendo en cuenta los criterios expuestos, se asume como **Problema científico**: ¿Qué alternativas metodológicas utilizar para la modelación de datos de brotes de dengue que permitan la estimación, predicción en tiempo real de parámetros epidemiológicos claves y la evaluación del impacto de intervenciones para el control de brotes?

Bajo las consideraciones antes planteadas, este trabajo de investigación tiene los siguientes objetivos:

Objetivo general

Proponer nuevas alternativas metodológicas para la modelación de datos en brotes de dengue que permitan la estimación y predicción en tiempo real de parámetros epidemiológicos claves.

Objetivos específicos

- Diseñar metodologías para la estimación y predicción en tiempo real de parámetros epidemiológicos de brotes de dengue de una y varias ondas epidémicas.
- Proponer diferentes enfoques metodológicos para la comparación de brotes epidémicos de dengue de una onda teniendo en cuenta la incertidumbre de los modelos en la estimación de los parámetros epidemiológicos.
- Aplicar los modelos no lineales con efectos mixtos en la modelación de datos de brotes de dengue para tener en cuenta la transmisión dentro de áreas en una misma región.
- Aplicar las metodologías propuestas a datos de brotes de dengue y Zika ocurridos en Cuba y Colombia.
- Confeccionar un paquete en R para la implementación de la metodología propuesta para brotes de una onda epidémica.

Novedad científica

En la presente tesis se proponen nuevas alternativas metodológicas para la modelación de brotes de una onda y multiondas que permiten la estimación y predicción en tiempo real de parámetros epidemiológicos claves. Es significativo que se aplican por primera vez los modelos mixtos no lineales que resultan un aporte científico para el estudio de la heterogeneidad de la transmisión entre diferentes áreas de una misma región. Por otra parte resulta también

un aporte, la exploración por primera vez de diferentes enfoques metodológicos para la comparación de brotes de una onda teniendo en cuenta la incertidumbre de los modelos en la estimación de parámetros epidemiológicos.

Otro aspecto novedoso de la tesis y que a su vez es también aporte científico de la misma, es la creación del paquete de R, *DengueRT*, que permite utilizar la metodología propuesta para brotes de una onda en el análisis en tiempo real de otros brotes en Cuba y la generalización de esta metodología a otros brotes de otras enfermedades, en otras condiciones epidemiológicas (por ejemplo, Zika en Colombia).

Estructura de la tesis

El cuerpo de la tesis se ha estructurado en cinco capítulos. El primer capítulo se ha dedicado a esbozar una breve introducción de la enfermedad, su transmisión y su distribución en Cuba y en el mundo. Así mismo, se hace un estudio de los modelos matemáticos aplicados al dengue, destacándose los modelo de transmisión de la enfermedad y los modelos fenomenológicos. Como la tesis está centrada en estos últimos modelos, se hace una revisión de los diferentes métodos que existen para la estimación de parámetros epidemiológicos en brotes de dengue, señalando sus principales limitaciones y alternativas.

En el segundo capítulo se presenta una metodología para modelar brotes de dengue de una onda y estimar parámetros epidemiológicos claves, como son, el tamaño final y acmé de la epidemia y realizar predicciones en tiempo real, utilizando varios modelos fenomenológicos no lineales donde los parámetros se estiman a través del promedio de modelos. La metodología propuesta es aplicada al brote de dengue ocurrido en dos áreas de salud del municipio Playa, La Habana, durante el brote de dengue del 2001/2002.

En el tercer capítulo se proponen diferentes métodos estadísticos para la estimación de parámetros epidemiológicos en brotes de una onda cuando los datos provienen de casos acumulados en dos o más regiones. En el caso de dos regiones, se proponen diferentes enfoques metodológicos (Frecuentistas y Bayesianos) que permiten tener en cuenta la incertidumbre de

modelos en la comparación de brotes. En el caso de más de dos regiones, se propone el uso de los modelos no lineales con efectos mixtos para tener en cuenta la heterogeneidad entre las áreas. La aplicación de estos métodos se muestra con la comparación de dos brotes de dengue en el municipio La Lisa en el 2006 y 2007 y a datos provenientes de siete áreas de salud del municipio Plaza del brote de dengue ocurrido en el 2006.

En el cuarto capítulo se abordan algunos métodos estadísticos para la modelación de brotes multiondas, es decir cuando hay más de una onda de infección. Se propone una variación del modelo de Richards multi-etapas, una nueva alternativa metodológica que utiliza la verosimilitud perfil. La aplicación de la propuesta se muestra usando datos de la provincia La Habana del brote de dengue del 2001/2002 y además se muestra su uso para evaluar el impacto de las intervenciones ocurridas para el control del brote.

En el quinto y último capítulo se presenta la implementación de los métodos, propuestos en el capítulo 2, en un paquete de R, *DengueRT*. El paquete *DengueRT* proporciona un conjunto de funciones para la estimación de parámetros y las predicciones en tiempo real de brotes de dengue de una onda, tiene en cuenta la incertidumbre de los modelos utilizando la metodología de promedio de modelo. Para una mejor comprensión del paquete, primeramente se hace una introducción, donde se presentan la implementación de todas las funciones (visibles o no para los usuarios) y los datos del paquete. También se describe cómo fue creado con el uso de **RStudio** y de la plataforma *R-forge* hasta su publicación en el CRAN (*Comprehensive R Archive Network*). La visualización de las funciones y gráficos del paquete se muestran para el brote de dengue ocurrido en dos áreas de salud del municipio Playa, La Habana, durante el brote de dengue del 2001/2002.

Capítulo 1

El dengue y la modelación de brotes de dengue

1.1 Introducción

En los últimos años ha existido un interés creciente en el uso de los modelos matemáticos para el análisis de brotes de enfermedades infecciosas [16,59], entre ellas el dengue. El dengue es una enfermedad infecciosa viral transmitida por mosquito que causa brotes epidémicos significativos, particularmente en área tropicales y subtropicales [13].

En este capítulo se hace una breve introducción a la enfermedad, describiendo la situación actual, la forma de transmisión, el agente causal de la enfermedad, factores de riesgos y la situación de dengue en Cuba. Así mismo se hace un estudio de los modelos matemáticos aplicados al dengue, destacándose los modelo de transmisión de la enfermedad y los modelos fenomenológicos.

Como esta investigación está centrada en los modelos fenomenológicos, se hace una revisión de los diferentes modelos que existen para la estimación de parámetros epidemiológicos en brotes de dengue, señalando sus principales limitaciones y alternativas. Finalmente, se hace una descripción de cada una de las bases de datos que serán utilizados para la aplicación de los métodos propuestos.

1.2 Dengue

El dengue es la infección viral más importante transmitida por artrópodos [32] que afecta al humano y ha emergido como un importante problema de salud en las regiones tropicales y subtropicales. Actualmente es endémica, con brotes epidémicos en América, el sudeste de Asia y el Pacífico [31]. Se estima que la incidencia global es de 390 millones (Intervalo de credibilidad 95%, 284-528) casos anualmente, de los cuales 50-100 millones son casos no severos y entre 250000-500000 son casos severos [7]. Esta enfermedad es causada por cuatro serotipos virales (DEN-1, DEN-2, DEN-3 y DEN-4) y es transmitida por la picada de un mosquito *Aedes aegypti* infectado. Los patrones de reproducción de la enfermedad en las poblaciones humanas están influenciados por una gran variedad de factores entre los que se encuentran el clima, factores ambientales y de urbanización, la densidad poblacional, problemas con el sistema de recolección de basura y el suministro de agua que favorecen la proliferación de sitios de cría del vector. Al no existir vacuna ni tratamiento específico para el dengue, Las acciones de control se basan fundamentalmente en el control del mosquito y en la vigilancia epidemiológica activa de la enfermedad y sus factores de riesgo.

La primera referencia de la circulación del dengue en Cuba data de 1782 en Remedios, pero ya existían reportes de enfermedades similares en ese mismo pueblo desde 1674. Coincidiendo con pandemias ocurridas en América, se reportaron brotes en 1827; 1848; 1849 y 1897. En el siglo XX el primer reporte apareció en 1944 en Ciudad de la Habana. Pero en 1977 ocurrió una nueva epidemia que afectó a todo el país en la que se reportaron 553 138 enfermos y se aisló el serotipo 1. En 1981, ocurrió la mayor epidemia de FHD en las Américas hasta nuestros días. En esta epidemia se detectaron 344 303 casos, de ellos 10 312 de FHD y 158 muertes y se aisló el serotipo 2 [32].

Después de 16 años sin transmisión, en 1997 se detectó en Santiago de Cuba un brote producido por DEN-2. Fueron confirmados 3 012 casos, de los cuales 205 se clasificaron como FHD y 12 fallecieron [32]. En el año 2000, en Ciudad de La Habana se reportó un brote de 138 casos confirmados asociados a los serotipos 3 y 4 que fue rápidamente controlado [75]. En Junio del 2001 una nueva epidemia azotó al país, confirmándose 12 889 nuevos casos, de ellos 78 fueron

de FHD y 3 fallecieron [62].

Cuba cuenta con un Sistema Nacional de Salud bien estructurado y cuyas estadísticas han sido citadas entre las de más alta calidad en el mundo. Presenta un programa de control de *Aedes aegypti* que monitorea y controla sistemáticamente las poblaciones de este mosquito, manteniéndolo en niveles comparativamente bajos con respecto al mundo. Sin embargo, a pesar de esto, sucesivos brotes han afectado al país en los últimos años en diferentes localidades debido a la presencia del vector, la alta susceptibilidad de la población y la afluencia constante de viajeros provenientes de áreas endémicas de dengue. La ocurrencia de la enfermedad en forma de brotes, con la circulación fundamentalmente de un solo serotipo, diferente en cada momento hace que nuestras condiciones epidemiológicas sean únicas. El reporte semanal de casos que se registra a los niveles de municipio y provincia constituyen la información principal para el desarrollo de los modelos matemáticos.

1.3 Modelos Matemáticos para brotes de dengue

Numerosos enfoques han sido utilizados para aclarar la epidemiología del dengue y del dengue hemorrágico. La epidemiología del dengue posee varias características ecológicas específicas, y por lo tanto, los determinantes epidemiológicos han sido explorado a menudo utilizando la modelación matemática.

1.3.1 Modelos de transmisión

Los modelos matemáticos de transmisión de la enfermedad pueden ser una forma poderosa de conocer mejor la dinámica de transmisión del dengue [8]. A lo largo de su historia, el modelación del dengue ha dependido en gran medida de la teoría originalmente desarrollada para representar la dinámica de la transmisión de la malaria: en particular, el modelo Ross-Mac Donald de transmisión de la malaria [81]. El modelo de Ross-Macdonald proporciona un marco básico para representar la interacción del patógeno con los mosquitos y humanos, u otros huéspedes vertebrados. Las muchas diferencias importantes entre el dengue y la epidemiología del paludismo han llevado a los investigadores a modificar este marco, pero la

estructura básica del modelo de Ross-Macdonald sigue siendo ampliamente aplicada en los modelos de transmisión del dengue.

Las características básicas de los modelos de transmisión del dengue son representaciones matemáticas de: (1) estructura de la población de humanos y su dinámica; (2) estructura de la población de vectores y su dinámica; (iii) la interacción entre estas dos especies; (iv) historia natural de las infecciones de dengue en mosquitos y humanos, incluyendo los posibles efectos de inmunidad; y (v) el impacto de las medidas de control [31].

Los modelos matemáticos de transmisión ponen énfasis en cada uno de los aspectos de la epidemiología del dengue y generalmente están diseñados para abordar preguntas específicas que pueden ser clasificadas en cinco categorías: (1) cuantificar la transmisión en diferentes entornos (por ejemplo, estimación del número reproductivo básico); (2) comprender los patrones temporales de la enfermedad (por ejemplo, caracterizar las fuerzas y mecanismos detrás de los patrones temporales observados); (3) predecir el patrón espacial de transmisión; (4) comprender la estructura serotipo / cepa; y (5) estimar el impacto potencial de los métodos de control [31].

Varias revisiones bibliográficas de los modelos de transmisión del dengue han aparecido recientemente en la literatura científica. Johansson y cols. [44] revisaron modelos que podrían ser usados para modelar el impacto de una vacuna contra el dengue con un enfoque particular en la representación de la respuesta inmune humana y su impacto en la transmisión. Ellos también revisaron enfoques que han sido utilizados para validar y parametrizar los modelos. Un parámetro clave en todos los enfoques fue el número reproductivo básico, R_0 , el número promedio de infecciones causadas por un individuo infeccioso cuando éste es introducido en una población de susceptibles.

R_0 intenta capturar la capacidad reproductiva de la enfermedad o su transmisibilidad. Para una enfermedad causada por un vector, R_0 puede ser entendido como el número de personas que serían infectadas a partir de una sola persona inicialmente infectada por un mosquito [13,56]. Este es un concepto clave en la epidemiología ya que a menudo sirve como parámetro

umbral que predice si una infección se propagará [17]. Cuando $R_0 < 1$ la infección desaparecerá en un corto período de tiempo. Pero cuando $R_0 > 1$ la infección será capaz de propagarse en la población. Altos valores de R_0 pueden indicar la posibilidad de una epidemia transcendental. Este parámetro también puede utilizarse para determinar la fracción crítica de vacunación. Johansson y cols. [44] también hicieron una revisión de los diferentes métodos que se utilizan en los modelos estudiados para estimar esta cantidad.

Andraud [5] realizó una revisión bibliográfica de los modelos determinísticos de la transmisión del dengue. Ellos clasificaron los modelos de interés publicados en modelos de un serotipo y modelos multi-serotipos. Para cada enfoque, se discutió la estructura fundamental, los supuestos de los parámetros, el comportamiento umbral y el impacto de las intervenciones. Reiner y cols. [69] hicieron una revisión bibliográfica que abarcaba los modelos de todos los patógenos transmitidos por mosquitos y los clasificaron según los supuestos biológicos que consideraron.

Más reciente, Chowell y cols. [15] realizaron una revisión bibliográfica de los progresos recientes en la modelación y caracterización de los patrones de crecimiento epidémico temprano a partir de datos de brotes de enfermedades infecciosas, y examinan los tipos de formulaciones matemáticas que son más utilizadas para captar una diversidad de perfiles de crecimiento epidémico temprano, que van desde dinámicas de crecimiento sub-exponenciales hasta exponenciales. Ellos también revisan modelos matemáticos que incorporan detalles espaciales o estructuras mixtas de poblaciones realistas, incluyendo modelos meta-poblacionales, modelos neuronales basados en individuos, y modelos sencillos tipo SIR que incorporan los efectos de cambios de comportamiento o mezclas no homogéneas.

A pesar de la popularidad en el uso de los modelos de transmisión del dengue, estos se ven obstaculizado por la notable incertidumbre sobre la epidemiología de la enfermedad, la dificultad de estar en un contexto de datos limitados y la ausencia de información confiable sobre los mecanismos de transmisión [14]. Una alternativa en esta situación es hacer uso de los modelos fenomenológicos.

1.3.2 Modelos fenomenológicos

Los modelos fenomenológicos (por ejemplo, Hsieh y cols. [38]) prescinden de los supuestos de los modelos matemáticos de transmisión. En su lugar, ellos hacen supuestos generales acerca de la forma de la curva de incidencia. En general, tales modelos fenomenológicos utilizan pocos parámetros, hacen suposiciones más sencillas, y son probables a dar estimaciones más robustas cuando son aplicados en contextos con datos muy limitados [54] (una situación muy común para epidemias históricas o cerca del inicio de una epidemia de una enfermedad emergente o re-emergente). Esta investigación se centra en el estudio de tales modelos para la estimación y predicción de parámetros epidemiológicos claves .

Entre los modelos fenomenológicos más utilizado para la modelación de datos de enfermedades infecciosas se encuentra el modelo de Richards [35–43,70,85], aunque recientemente Chowell y cols. [14–17] han utilizado una generalización del modelo de Richards, el modelo de Richards generalizado, que ha ayudado a caracterizar los perfiles de crecimiento epidémico temprano a partir de datos de brotes de enfermedades infecciosas como Zika, Ébola y VIH SIDA. El modelo de Richards y esta generalización considera sólo el tamaño de la población infecciosa acumulada con saturaciones en el crecimiento cuando el brote avanza, posiblemente causadas por la implementación de las medidas de control.

La premisa básica del modelo de Richards es que la curva de incidencia está compuesta de un único pico de alta incidencia, el acmé de la epidemia [35,42]. El acmé de la epidemia es el punto en el tiempo, donde la razón de acumulación cambia de un crecimiento a un decrecimiento. Puede ser determinado con precisión encontrando el punto de inflexión de la curva epidémica de casos acumulados. Cuando hay más de una onda epidémica al punto del tiempo que indica el fin del crecimiento exponencial inicial se le llamará acmé de la epidemia pero al punto en el tiempo donde la razón de crecimiento del número de casos acumulados comienza a incrementarse nuevamente, significando el inicio de una nueva onda se le llamará punto de retorno [35,38–40]. El acmé de la epidemia y los puntos de retorno nos permite estudiar el impacto de las medidas de intervención.

El modelo de Richards se ha utilizado con frecuencia para modelar los casos notificados acu-

mulados en los brotes de dengue. En particular, Hsieh y cols. [35, 42] propuso utilizar el modelo de Richards para estimar el acmé de la epidemia, el número reproductivo básico R_0 y el tamaño final de la epidemia durante un brote de una onda. El tamaño final de la epidemia o número máximo de casos es una de las propiedades asintóticas más importantes en los modelos epidemiológicos y se define como el número total de individuos infectados en una epidemia. Es una medida cuantitativa muy importante porque se relaciona de forma estrecha con los costos de la epidemia [3].

Uno de los parámetros del modelo de Richards es la razón de crecimiento exponencial inicial, la cual es una importante medida de la propagación de la enfermedad, y es comúnmente utilizada para inferir el número reproductivo básico R_0 . R_0 está determinado por la razón de crecimiento inicial y la distribución del tiempo de generación, es decir, el tiempo de intervalo promedio entre la infección del caso primario y los casos secundarios que son el resultado de él. Si se supone una infección donde todos los tiempos de todas las infecciones secundarias son exactamente igual al intervalo de generación T , tal distribución conforma la llamada distribución Delta. Se ha demostrado matemáticamente que, dado una razón de crecimiento γ , la expresión $R_0 = e^{\gamma \cdot T}$ proporciona una cota superior del número reproductivo básico indiferente de la distribución del intervalo de generación utilizado [84].

En el caso del dengue, para tener en cuenta los períodos de la incubación extrínsecos e intrínsecos así como duración de viraemia, se utiliza para la estimación del número reproductivo básico R_0 un tiempo de generación estimado de $T = 24$ días o de $T = \frac{24}{7}$ semanas con un rango de 16 – 34 días [38]. El error estándar aproximado para la estimación de R_0 utilizando el método delta está dado por la expresión $SE(R_0) \approx T \cdot e^{\gamma \cdot T} SE(\gamma)$.

Hsieh y cols. [35, 42] también demostraron que el modelo de Richards se puede utilizar para la predicción en tiempo real de la severidad del brote mediante la estimación de la capacidad de carga (es decir, el tamaño final de la epidemia), así como la detección en tiempo real del acmé de la epidemia. Una predicción en tiempo real es un procedimiento en el cual tanto el tamaño final y el acmé de la epidemia son estimados tan temprano como sea posible.

Una gran variedad de modelos de crecimiento no lineales se han desarrollado para modelar

los datos de crecimiento. Entre ellos, están el modelo logístico de 3 parámetros (logístico 3P) [27, 35, 83], el modelo logístico de 5 parámetros (logístico 5P) [50], el modelo Sigmoideo Emax [55], el modelo de Gompertz [27, 83] y el modelo de Weibull [77]. Todos estos modelos podrían ajustar bien los datos epidémicos, por lo que aparecería un aspecto fundamental en la modelación estadística, la selección de modelos.

Típicamente la selección del modelo de mejor ajuste dentro de todos los modelos posibles ignora la incertidumbre de la selección de modelos en la estimación e inferencia. Por estas razones, varios autores, entre los cuales están Burnham & Anderson [10], Posada [65], Moon [57], Claeskens & Hjort [18] y Lin [51] abogaron por el uso de técnicas de promedio de modelos (*model averaging*) para realizar inferencias multi-modelos en los parámetros estimados. El promedio de modelos es un método que tiene en cuenta todos los modelos ajustados para la estimación de los parámetros de interés. Está basado en un promedio ponderado de los parámetros estimados obtenido de los diferentes modelos, dándoles mayor peso a esos modelos que mejor ajustan los datos [26, 45, 58]. Esta podría ser una alternativa para la estimación y predicción en tiempo real de parámetros epidemiológicos en brotes de dengue de una onda e inclusive para la comparación de brotes a través de estos parámetros, aspecto poco abordado en la literatura.

En ocasiones el número de casos acumulados reportados se colecciona para todas las áreas urbanas en una región en particular. Muchas veces resulta de interés estimar parámetros epidemiológicos claves o estudiar la heterogeneidad o variabilidad regional de la transmisión de la enfermedad a través del número reproductivo básico R_0 . La modelación, para estimar parámetros epidemiológicos, se lleva a cabo usualmente para cada área por separado, ajustando el modelo de Richards individual para cada área específica. Este abordaje es útil cuando se está interesado en modelar el comportamiento de un área en particular o conjunto fijo de áreas, pero no es adecuado cuando las áreas son consideradas como muestra de una población (quizás hipotética) y la inferencia debe centrarse en esta población. Además, con este abordaje se ignora la variabilidad entre las áreas.

Una forma eficaz de abordar la modelación de datos de brotes de dengue cuando interesa

tener en cuenta la estructura urbana de la población en una región, es decir la transmisión entre áreas en una región determinada o estimar el comportamiento promedio de una área, y la variabilidad entre y dentro de las áreas, sería utilizando un modelo no lineal con efectos mixtos, no utilizados hasta el momento para la modelación de este tipo de datos en brotes de dengue.

En la modelación de datos de brotes de dengue cuando hay más de una onda de infección, se ha propuesto el empleo de una variación del modelo de Richards: El modelo de Richards multi-etapas, el cual hace distinción de dos puntos importantes en el tiempo, el acmé de la epidemia y los puntos de retorno. Recientemente, Hsieh [35, 38–40] demostró su utilización para modelar brotes multiondas de SARS, dengue, Zika, pH1N1 y Ébola con el fin de estimar los parámetros epidemiológicos claves.

El procedimiento del modelo de Richards multi-etapas, utilizado por Hsieh [35, 38–40] requiere de 5 pasos:

1. Ajustar el modelo de Richards al número de casos acumulados en semanas sucesivas utilizando una rutina de mínimos cuadrados estándar. Para brotes de una sola onda, la estimación de los parámetros $\theta = (\alpha, \gamma, k, \eta)$ convergerá en cuanto la trayectoria se aproxime a la capacidad de carga α .
2. Si los parámetros estimados permanecen convergente hasta que no sean detectados más nuevos casos, el brote tiene solo una onda. Sin embargo, si las estimaciones comienzan a divergir desde una valor fijo, uno sabe que ha ocurrido el punto de retorno que denota el inicio de una segunda fase.
3. Localizar el punto de retorno, t_{min} , que separa las dos fases en forma de S de una epidemia como el mínimo local de la curva de incidencia.
4. Ajustar el modelo de Richards a la curva de casos acumulados nuevamente, pero comenzando desde $t_{min} + 1$, la unidad de tiempo después del inicio de la segunda fase. Los parámetros estimados $\theta = (\alpha, \gamma, k, \eta)$ convergerán nuevamente en cuanto la curva se aproxime a la capacidad de carga para la segunda fase.

5. Repetir los pasos 2-4 en los eventos que ocurran más fases hasta que termine el brote.

El procedimiento del modelo de Richards multi-etapas propuesto por Hsieh [35, 38–40] considera los segmentos sucesivos de casos acumulados en forma de S de forma separada, constituyendo esto desde nuestro punto de vista una limitación. Una alternativa a esta limitación podría ser proponer una metodología que permita la estimación de los parámetros epidemiológicos teniendo en cuenta los segmentos de la curva epidémica simultáneamente y haciendo uso de otros modelos no lineales para realizar inferencias multimodelos.

Otro aspecto revisado en la literatura relacionado con los modelos fenomenológicos, fue el método de estimación llevado a cabo para estimar los parámetros. En los trabajos de Hsieh [35–38, 40–43], que utilizan el modelo de Richards, los parámetros fueron estimados con herramientas de aproximación mínimo cuadrática utilizando los software SAS o Matlab y los intervalos de confianza de los parámetros fueron obtenidos empíricamente utilizando subrutinas de estimación mínimo cuadrática. La estadística inferencial formal del ajuste mínimo cuadrático está basada en asumir errores independientes, distribuidos normalmente con varianza constante.

En los trabajos de Ma y cols. [54], los parámetros fueron estimados utilizando el método de máxima verosimilitud y asumiendo que la incidencia durante cada periodo de reportes sigue una distribución de Poisson. Ellos además, utilizaron este enfoque para estimar los intervalos de confianza de los parámetros a través de la verosimilitud perfil. Finalmente, en los trabajos de Chowell y cols. [14–17] las estimaciones de los parámetros del modelo fueron obtenidas ajustando la curvas por mínimos cuadrados no lineales utilizando el algoritmo de Levenberg-Marquardt en el MATLAB. La incertidumbre de los parámetros se cuantificó simulado 200 realizaciones estocásticas de las curvas de mejor ajuste usando bootstrap paramétrico, con estructura de error Poisson. Las curvas epidémicas simuladas fueron utilizadas para obtener los intervalos de confianza del 95% para cada parámetro.

A pesar que en la literatura revisada no se encontraron métodos de estimación de los parámetros de los modelos fenomenológicos por el enfoque Bayesiano aplicados a brotes de dengue, en los

trabajos de Firat y cols. [27] se discute el enfoque Bayesiano como una alternativa al análisis clásico de modelos no lineales para datos de crecimiento de codornices japonesas y se comparan las estimaciones de los parámetros de tres modelos no lineales: Gompertz, Richards y logístico; utilizando ambos enfoques. Por el enfoque Frecuentista se aplica el método de los mínimos cuadrado no lineales utilizando el método iterativo de Gauss-Newton con la ayuda del software SAS. Por el enfoque Bayesiano, las muestras a posteriori de los parámetros del modelo requieren de técnicas de muestreo por el método de Monte Carlo con cadenas de Markov y fueron obtenidas por el programa WinBUGS. Se asumió distribución a priori Normal para todos los parámetros del modelo y distribución a priori no informativa para la varianza de los residuos.

1.4 Descripción de los datos en la modelación

En la presente tesis se utilizaron datos de los casos de dengue reportados en los brotes ocurridos en la Habana en 2001-2002 y en 2006-2007. La aplicación de cada modelo requirió la selección de un conjunto de datos diferentes. Para todos los modelos, los casos confirmados se agruparon por semanas epidemiológicas de acuerdo a la fecha de primeros síntomas.

Para la modelación de brotes de una onda teniendo en cuenta la incertidumbre de modelos utilizando varios modelos no lineales se seleccionaron los datos del brote de 2001/2002, en las áreas de salud "26 de Julio" y "Jorge Ruiz Ramirez" del municipio Playa (Figura 5.4, Anexo 2). Estos datos también fueron utilizados para mostrar el uso del paquete de R, *DengueRT*

Para explorar diferentes enfoques metodológicos para la modelación teniendo en cuenta la incertidumbre en la comparación de brotes se utilizaron los datos de casos reportados de los brotes del 2006 y 2007 ocurridos en el municipio La Lisa (Figura 5.5, Anexo 2).

Los datos de casos reportados acumulados durante el brote del 2006 en las siete áreas de salud del municipio Plaza, enumeradas como: 1 ("15 y 18"), 2 ("19 de Abril"), 3 ("Corynthia"), 4 ("Moncada"), 5 ("Puentes Grandes"), 6 ("Plaza") y 7 ("Rampa"), fueron utilizados para ilustrar la modelación de la heterogeneidad de la transmisión entre áreas de una misma región

(Figura 5.6, Anexo 2).

Finalmente, se utilizaron los datos de los casos reportados para toda la provincia La Habana en el brote de 2001/2002 para ilustrar una nueva alternativa metodológica para la modelación de brotes multiondas (Figura 5.7, Anexo 2).

1.5 Conclusiones parciales

El dengue ha emergido como un importante problema de salud a escala mundial. Los determinantes epidemiológicos del dengue han sido explorados con frecuencia utilizando modelos matemáticos. Dentro de los modelos matemáticos más utilizados están los modelos matemáticos de transmisión y los modelos fenomenológicos.

A pesar de la popularidad en el uso de los modelos de transmisión del dengue, estos se ven obstaculizados por la notable incertidumbre en los parámetros epidemiológicos y por la dificultad de estar en un contexto de datos limitados. Una alternativa en esta situación es hacer uso de los modelos fenomenológicos, los cuales hacen supuestos generales acerca de la forma de la curva de incidencia, utilizan pocos parámetros y tienen la ventaja que solo necesita de los datos de incidencia o los datos de casos acumulados.

Existen diferentes métodos estadísticos que hacen uso de los modelos fenomenológicos para la estimación de parámetros epidemiológicos claves en brotes de dengue. En el capítulo se hace una revisión de estos métodos señalando sus principales limitaciones y alternativas.

Capítulo 2

Incertidumbre de modelos de brotes de una onda

2.1 Introducción

El desarrollo de herramientas estadísticas para la estimación y predicción de parámetros en epidemias de enfermedades infecciosas emergentes y reemergentes representa una prioridad para las autoridades de salud en su trabajo para prevenir y mitigar las amenazas de enfermedades [79]. Entre los modelos fenomenológicos más utilizados para la modelación de datos de enfermedades infecciosas se encuentra el modelo de Richards. En particular, Hsieh y cols. [35, 42] propuso utilizar este modelo para estimar y realizar predicciones en tiempo real del acmé, tamaño final de la epidemia y el número reproductivo básico R_0 en brotes de dengue de una onda.

Sin embargo, una gran variedad de modelos fenomenológicos no lineales se han desarrollado para modelar los datos de crecimiento, los cuales podrían ajustar bien los datos de brotes epidémicos de dengue. Usualmente, uno sigue el procedimiento de estimación llamado post-selección, es decir, selecciona el modelo de mejor ajuste de un conjunto de modelos candidatos e ignora la incertidumbre de la selección del modelo en la estimación y en la inferencia. Por esta razón, varios autores han abogado por estimar los parámetros primarios en el modelo utilizando la técnica de promedio de modelos (*model averaging*) y realizar inferencia multi-modelos.

En este capítulo se propone una metodología para analizar los datos de un brote de dengue de una onda y estimar el tamaño final de la epidemia, el punto de retorno y realizar predicciones en tiempo real utilizando varios modelos en el que los parámetros se estiman a través del promedio de modelos. El método propuesto se aplica al brote de dengue que ocurrió en dos áreas de salud del municipio Playa en La Habana, durante el brote de dengue del 2001/2002.

2.2 Modelos no lineales y métodos de estimación

Sea Y_t el número de casos acumulados notificados en el tiempo t . En esta investigación en un primer escenario se asume que Y_t tiene distribución normal asintótica con media $\mu(t, \theta)$ y varianza σ^2 , $Y_t \sim N(\mu(t, \theta), \sigma^2)$ y en un segundo escenario, y a modo de comparación, se asume que $Y_t \sim Poisson(\mu(t, \theta))$. Aquí, $\mu(t, \theta)$ es la estructura de la media o el modelo no lineal conocido a utilizar y θ el vector de parámetros a ser estimados.

Asumir que la variable Y_t tiene distribución normal asintótica implicará que los valores predichos por el modelo no sean enteros, pero en nuestro caso no es un problema ya que estos pueden ser aproximados al entero más cercano. Además, como los datos están cerca de cero solo al inicio del brote y estos aumentan con el tiempo por ser datos acumulados; asumir Y_t con distribución normal asintótica tampoco traerá problemas ya que solo se podrían obtener valores predichos negativos al inicio de la epidemia, lo que en nuestra situación no tiene grandes consecuencias. Teóricamente, este supuesto está basado en que la distribución límite de una distribución de Poisson ($Poisson(\lambda)$) cuando $\lambda \rightarrow \infty$ es la distribución normal. De todas formas siempre será necesario chequear los supuestos del modelo.

El modelo de Richards pertenece a la familia de modelos no lineales. Su expresión analítica aparece en la tabla 2.1 y el vector de parámetros a ser estimado, como se ha señalado por Hsieh y cols. [35, 42], es $\theta = (\alpha, \gamma, k, \eta)$, donde α es el tamaño final de la epidemia, γ es la tasa de crecimiento intrínseca per cápita de la población infectada, k es el exponente de la desviación de la curva logística estándar y η es el acmé de la epidemia.

Además del modelo de Richards, la Tabla 2.1 presenta un conjunto de cinco modelos no

lineales. Note que todos estos modelos están en escala del tiempo calendario (para $t = 1$ es la primera semana epidemiológica en la que ocurrió el brote). El modelo logístico de tres parámetros (logístico 3P) es un caso especial del modelo de Richards cuando el exponente $k = 1$. En el modelo logístico 3P, la curva de crecimiento es simétrica alrededor del punto de máxima tasa de crecimiento (acmé de la epidemia) y tiene períodos iguales de crecimiento lento y rápido. El modelo de Gompertz es otro caso especial del modelo de Richards cuando $\gamma \rightarrow 0$ y se utiliza con frecuencia en situaciones donde el crecimiento no es simétrico alrededor del punto de inflexión. Hay muchas variantes del modelo de Weibull, la que utilizamos en este trabajo es una modificación del modelo de Gompertz cuando su variable independiente está en escala logarítmica.

Los modelos Sigmoido Emax y logístico de cinco parámetro (logístico 5P) son comúnmente utilizados en los modelos de dosis-respuesta. El modelo Sigmoido Emax se obtiene por una transformación matemática del modelo logístico 3P y rescalando la variable independiente por una transformación logarítmica. El modelo logístico 5P también es obtenido rescalando la variable independiente a través de una transformación logarítmica y haciendo una reparametrización, de modo que el modelo evaluado en su punto de inflexión (η) alcance el 50% de la respuesta máxima.

Note que para todos los modelos en la tabla 2.1 el acmé (η) y el tamaño final de la epidemia α son parámetros en el modelo y que la primera derivada de $\mu(t, \theta)$, $\mu'(t, \theta) = \frac{\partial \mu(t, \theta)}{\partial t}$, es el número de casos en el tiempo t , es decir, la incidencia en el tiempo t . Excepto para el modelo logístico 3P, $\mu'(t, \theta)$ no es simétrica respecto al acmé de la epidemia.

2.2.1 Métodos de estimación

Existen varios métodos para estimar el vector de parámetros en los modelo no lineal presentados previamente. En este epígrafe se presentan algunos de estos procedimientos: Mínimos cuadrados no lineales con los métodos iterativos *Gauss-Newton* y *Newton*, mínimos cuadrados generalizados y el método de máxima verosimilitud para datos normales y no normales [77].

Suponga que se tienen n observaciones $(x_i, y_i), i = 1, 2, \dots, n$, el modelo no lineal con relación

Tabla 2.1: Modelos no lineales propuestos para el ajuste de brotes de dengue.

Modelos	$\mu(t, \theta)$	$\mu'(t, \theta)$
Richards	$\frac{\alpha}{[1+k \cdot e^{-\gamma \cdot k(t-n)}]^{\frac{1}{k}}}$	$\gamma \mu(t) \left[1 - \left(\frac{\mu(t)}{\alpha} \right)^k \right]$
Logístico 3P	$\frac{\alpha}{1+e^{-\gamma(t-n)}}$	$\gamma \mu(t) \left[1 - \frac{\mu(t)}{\alpha} \right]$
Logístico 5P	$\alpha + \frac{\alpha_0 - \alpha}{[1+(2^{\frac{1}{k}} - 1)(\frac{t}{\eta})^\gamma]^k}$	$-\frac{k\gamma}{t} [\mu(t) - \alpha] \left[1 - \left(\frac{\mu(t) - \alpha}{\alpha_0 - \alpha} \right)^{\frac{1}{k}} \right]$
Sigmoideo Emax	$\alpha_0 + \frac{t^n(\alpha - \alpha_0)}{t^n + \eta^n}$	$\frac{n}{t} [\mu(t) - \alpha_0] \left[1 - \frac{\mu(t) - \alpha_0}{\alpha - \alpha_0} \right]$
Gompertz 4P	$\alpha_0 + (\alpha - \alpha_0)e^{-e^{-\gamma(t-n)}}$	$-\gamma [\mu(t) - \alpha_0] \ln \left[\frac{\mu(t) - \alpha_0}{\alpha - \alpha_0} \right]$
Weibull	$\alpha + (\alpha_0 - \alpha)e^{-\left(\frac{t}{\eta}\right)^\gamma}$	$\frac{\gamma}{t} [\mu(t) - \alpha] \ln \left[\frac{\mu(t) - \alpha}{\alpha_0 - \alpha} \right]$

funcional f conocida es de la forma:

$$y_i = f(x_i, \theta^*) + \epsilon_i \quad (2.1)$$

donde $E(\epsilon_i) = 0$ para $1 \leq i \leq n$, y el verdadero valor θ^* de θ , $\theta^* \in \Theta \subseteq \mathbf{R}^p$

Mínimos cuadrados no lineales

El estimador mínimo cuadrático $\hat{\theta}$, se obtiene al minimizar la suma de cuadrados:

$$S(\theta) = \sum_{i=1}^n (y_i - f(x_i, \theta))^2 = \|\mathbf{y} - \mathbf{f}(\theta)\|^2 \quad \theta \in \Theta \subseteq \mathbf{R}^p \quad (2.2)$$

donde

$$\mathbf{y} = (y_1, \dots, y_n)^T$$

$$f_i(\theta) = f(x_i, \theta)$$

$$\mathbf{f}(\theta) = (f_1(\theta), f_2(\theta), \dots, f_n(\theta))^T$$

Bajo condiciones de regularidad, que incluyen que los errores ϵ_i , $1 \leq i \leq n$ sean independiente e idénticamente distribuidos con varianza σ^2 , tenemos que $\hat{\theta}$ y $s^2 = \frac{S(\hat{\theta})}{(n-p)}$ son estimadores consistentes de θ^* y σ^2 , respectivamente.

Si se denota

$$\mathbf{F}_\bullet(\theta) = \frac{\partial \mathbf{f}(\theta)}{\partial \theta^T} = \left[\left(\frac{\partial f_i(\theta)}{\partial \theta_j} \right) \right] \quad (2.3)$$

y

$$\mathbf{F}_\bullet = \mathbf{F}_\bullet(\theta^*) \quad \hat{\mathbf{F}}_\bullet = \mathbf{F}_\bullet(\hat{\theta}) \quad (2.4)$$

cuando $f(x_i, \theta)$ es diferenciable con respecto a θ , tenemos las ecuaciones

$$\frac{\partial S(\theta)}{\partial \theta_r} \Big|_{\theta=\hat{\theta}} = \sum_{i=1}^n 2(y_i - f_i(\theta))(-1) \frac{\partial f_i(\theta)}{\partial \theta_r} \Big|_{\theta=\hat{\theta}} = 0 \quad r = 1, \dots, p \quad (2.5)$$

lo que equivale a tener la ecuación

$$0 = \hat{\mathbf{F}}_\bullet^T (\mathbf{y} - \mathbf{f}(\hat{\theta})) = \hat{\mathbf{F}}_\bullet^T \hat{\boldsymbol{\epsilon}} \quad (2.6)$$

Igual que en la teoría de los modelos lineales, si $\hat{\mathbf{P}}_{\mathbf{F}} = \hat{\mathbf{F}}_\bullet (\hat{\mathbf{F}}_\bullet^T \hat{\mathbf{F}}_\bullet)^{-1} \hat{\mathbf{F}}_\bullet^T$ es la matriz idempotente que proyecta ortogonalmente \mathbb{R}^n sobre el espacio columna de $\hat{\mathbf{F}}_\bullet$, la ecuación anterior deriva en la ecuación normal para el modelo no lineal

$$\hat{\mathbf{P}}_{\mathbf{F}} \hat{\boldsymbol{\epsilon}} = 0 \quad (2.7)$$

Al ser $f(x, \theta)$ una función no lineal en θ , en la mayoría de los casos no pueden ser resueltos analíticamente. Esto obliga al uso de métodos iterativos para la solución del problema. Los procedimientos más utilizados son el de *Gauss-Newton* y el de *Newton*.

Método de Gauss-Newton. Este procedimiento se basa en la aproximación lineal de la función mediante un desarrollo de Taylor de primer orden de la función f .

Si se denota $\theta^{(a)}$ a una aproximación del estimador de mínimos cuadrados $\hat{\theta}$ del modelo no lineal 2.1. Para θ cercano a $\theta^{(a)}$ sabemos que es válida la aproximación lineal de Taylor

$$\mathbf{f}(\theta) \approx \mathbf{f}(\theta^{(a)}) + \mathbf{F}_\bullet^{(a)} (\theta - \theta^{(a)}) \quad (2.8)$$

donde $\mathbf{F}_\bullet^{(a)} = \mathbf{F}_\bullet(\theta^{(a)})$.

Así, al vector de residuos $\mathbf{r}(\theta) = \mathbf{y} - \mathbf{f}(\theta)$ lo podemos aproximar por $\mathbf{r}(\theta^{(a)}) - \mathbf{F}_\bullet^{(a)} (\theta - \theta^{(a)})$, que sugiere la aproximación

$$S(\theta) \approx \mathbf{r}(\theta^{(a)})^T \mathbf{r}(\theta^{(a)}) - 2\mathbf{r}(\theta^{(a)})^T \mathbf{F}_\bullet^{(a)} (\theta - \theta^{(a)}) + (\theta - \theta^{(a)})^T \mathbf{F}_\bullet^{(a)T} \mathbf{F}_\bullet^{(a)} (\theta - \theta^{(a)}) \quad (2.9)$$

El lado derecho tiene un mínimo con respecto a θ cuando

$$\theta - \theta^{(a)} = (\mathbf{F}_{\bullet}^{(a)T} \mathbf{F}_{\bullet}^{(a)})^{-1} \mathbf{F}_{\bullet}^{(a)T} \mathbf{r}(\theta^{(a)}) = \delta^{(a)} \quad (2.10)$$

Así, dada una aproximación $\theta^{(a)}$, la siguiente podría ser

$$\theta^{(a+1)} = \theta^{(a)} + \delta^{(a)} \quad (2.11)$$

Este procedimiento es el que se conoce como el método de *Gauss-Newton*. Para $\theta^{(1)}$ cercano a θ^* y el tamaño de la muestra suficientemente grande, tenemos que $\theta^{(a)} \rightarrow \hat{\theta}$ cuando $a \rightarrow \infty$.

Método de Newton. Este es un método más general que el de *Gauss-Newton* y se basa en el desarrollo de Taylor de segundo orden de $S(\theta)$

Si se denota

$$\mathbf{g}(\theta) = \frac{\partial S(\theta)}{\partial \theta} \quad (2.12)$$

al gradiente de $S(\theta)$ y

$$\mathbf{H}(\theta) = \frac{\partial^2 S(\theta)}{\partial \theta \partial \theta^T} \quad (2.13)$$

al Hesiano de $S(\theta)$, tenemos la aproximación cuadrática

$$S(\theta) \approx S(\theta^{(a)}) + \mathbf{g}(\theta^{(a)})^T (\theta - \theta^{(a)}) + \frac{1}{2} (\theta - \theta^{(a)})^T \mathbf{H}(\theta^{(a)}) (\theta - \theta^{(a)}) \quad (2.14)$$

Como

$$\frac{\partial^2 S(\theta)}{\partial \theta_r \partial \theta_s} = 2 \sum_{i=1}^n \left\{ \frac{\partial f_i(\theta)}{\partial \theta_r} \frac{\partial f_i(\theta)}{\partial \theta_s} - (y_i - f_i(\theta)) \frac{\partial^2 f_i(\theta)}{\partial \theta_r \partial \theta_s} \right\}, \quad (2.15)$$

Si se utiliza como aproximación de $\mathbf{H}(\theta^{(a)})$ su valor esperado en θ

$$\mathbb{E} \left(\frac{\partial^2 S(\theta)}{\partial \theta \partial \theta^T} \right) = 2 \mathbf{F}_{\bullet}^T(\theta) \mathbf{F}_{\bullet}(\theta) \quad (2.16)$$

El mínimo con respecto a θ en la aproximación cuadrática (2.14) ocurre cuando

$$\theta - \theta^{(a)} = - \left[\mathbf{H}(\theta^{(a)}) \right]^{-1} \mathbf{g}(\theta^{(a)}) = - \left[\mathbf{H}^{-1} \mathbf{g} \right]_{\theta=\theta^{(a)}} \quad (2.17)$$

Así, el paso es $\delta^{(a)} = - \left[\mathbf{H}^{-1} \mathbf{g} \right]_{\theta=\theta^{(a)}}$

Mínimos cuadrados generalizados

Este método se utiliza para resolver el problema de regresión no lineal cuando se tiene $Var(\epsilon) = \sigma^2 \mathbf{V}$, donde \mathbf{V} es una matriz definida positiva conocida.

En esta situación, se tiene que minimizar la suma de cuadrados

$$S_{MCG}(\theta) = [\mathbf{y} - \mathbf{f}(\theta)]^T \mathbf{V}^{-1} [\mathbf{y} - \mathbf{f}(\theta)] \quad (2.18)$$

Usando la descomposición de Cholesky de la matriz \mathbf{V} , se puede transformar el modelo no lineal en un problema de mínimos cuadrados ordinarios (MCO).

Si $\mathbf{V} = \mathbf{U}^T \mathbf{U}$ es la descomposición de Cholesky de \mathbf{V} , con \mathbf{U} una matriz triangular superior, llamando \mathbf{R} a $(\mathbf{U}^T)^{-1}$, se tienen que

$$\mathbf{R}\mathbf{y} = \mathbf{R}\mathbf{f}(\theta) + \mathbf{R}\epsilon \quad (2.19)$$

Si $\mathbf{Z} = \mathbf{R}\mathbf{y}$, $\mathbf{K}(\theta) = \mathbf{R}\mathbf{f}(\theta)$ y $\eta = \mathbf{R}(\epsilon)$, se tiene

$$\mathbf{Z} = \mathbf{K}(\theta) + \eta \quad (2.20)$$

Con $\mathbb{E}(\eta) = \mathbf{R}\mathbb{E}(\epsilon) = 0$ y $Var[\eta] = \sigma^2 \mathbf{R}\mathbf{V}\mathbf{R}^T = \sigma^2 \mathbf{I}_n$

Así, la suma de cuadrados del método de mínimos cuadrados generalizados (MCG), es la misma que la del método de mínimos cuadrados ordinarios (MCO) del problema transformado.

$$S_{MCG}(\theta) = [\mathbf{y} - \mathbf{f}(\theta)]^T \mathbf{V}^{-1} [\mathbf{y} - \mathbf{f}(\theta)] = [\mathbf{Z} - \mathbf{K}(\theta)]^T [\mathbf{Z} - \mathbf{K}(\theta)] \quad (2.21)$$

y el estimador MCO ($\hat{\theta}_{MCO}$) para el modelo transformado, es el estimador MCG ($\hat{\theta}_{MCG}$) del modelo sin transformar.

Si se denota $\mathbf{K}_{\bullet}(\theta) = \frac{\partial \mathbf{K}(\theta)}{\partial \theta^T} = \mathbf{R} \frac{\partial \mathbf{f}(\theta)}{\partial \theta^T} = \mathbf{R}\mathbf{F}_{\bullet}(\theta)$, una aproximación para el valor mínimo del problema de mínimos cuadrados generalizados se obtiene entonces mediante

$$\theta - \theta^{(a)} = (\mathbf{F}_{\bullet}^{(a)T} \mathbf{V}^{-1} \mathbf{F}_{\bullet}^{(a)})^{-1} \mathbf{F}_{\bullet}^{(a)T} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{f}(\theta^{(a)})) = \delta^{(a)} \quad (2.22)$$

Esto se lleva al método iterativo $\theta^{(a+1)} = \theta^{(a)} + \delta^{(a)}$ donde

$$\begin{aligned}
\delta^{(a)} &= (\mathbf{F}_{\bullet}^{(a)T} \mathbf{V}^{-1} \mathbf{F}_{\bullet}^{(a)})^{-1} \mathbf{F}_{\bullet}^{(a)T} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{f}(\theta^{(a)})) \\
&= \left(\mathbf{K}_{\bullet}^{(a)T} \mathbf{K}_{\bullet}^{(a)} \right)^{-1} \mathbf{K}_{\bullet}^{(a)T} (\mathbf{Z} - \mathbf{K}(\theta^{(a)}))
\end{aligned} \tag{2.23}$$

donde la última ecuación es el método de *Gauss-Newton* para el modelo transformado.

Estimadores de máxima verosimilitud

El estimador de máxima verosimilitud, $\hat{\theta}_{EMV}$, de θ puede obtenerse maximizando la función de verosimilitud.

Suponga que los errores ϵ_i son independientes e igualmente distribuidos, *i.i.d.* y que tienen distribución $N(0, \sigma^2)$, entonces se tiene que la función de verosimilitud es

$$L(\mathbf{y} | \theta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - f(x_i, \theta))^2}{\sigma^2}\right) \tag{2.24}$$

y la función log-verosimilitud está dada por

$$\begin{aligned}
l(\theta, \sigma^2) &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - f(x_i, \theta))^2 \\
&= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} S(\theta)
\end{aligned} \tag{2.25}$$

Dado σ^2 , La log-verosimilitud es máxima con respecto a θ cuando $S(\theta)$ es mínimo. O sea, que el estimador de máxima verosimilitud de θ resulta ser el estimador de mínimos cuadrados no lineales, $\hat{\theta}_{EMV} = \hat{\theta}_{MCO}$

Dado θ , se tiene que $\frac{\partial l(\theta, \sigma^2)}{\partial \sigma^2} = 0$ cuando $\sigma^2 = \frac{S(\theta)}{n}$. Luego, $\hat{\sigma}^2 = \frac{S(\hat{\theta}_{EMV})}{n}$ y $\hat{\theta}_{EMV}$ maximizan la función de log-verosimilitud $l(\theta, \sigma^2)$ y bajo condiciones apropiadas de regularidad de la función f , $\hat{\theta}_{EMV}$ es asintóticamente eficiente. Cuando los datos no son normales, el estimador de máxima verosimilitud no coincide con el de los mínimos cuadrados por la dependencia de la función de verosimilitud L de $\mathbf{f}(\theta)$, es decir $L(\mathbf{f}(\theta))$.

En este caso, se supone que el modelo es suficientemente regular como para que el EMV de $\hat{\theta}_{EMV}$ resulte ser solución de

$$0 = \frac{\partial L}{\partial \theta} = \left(\frac{\partial \mathbf{f}}{\partial \theta^T} \right)^T \frac{\partial L}{\partial \mathbf{f}} \tag{2.26}$$

Si $\theta^{(a)}$ es la a -ésima aproximación de $\hat{\theta}_{EMV}$, dado el desarrollo de Taylor de $\frac{\partial L}{\partial \theta}$ en $\hat{\theta}_{EMV}$, tenemos que la estimación $\theta^{(a+1)} = \theta^{(a)} + \delta^{(a)}$ esta dada por

$$\left[\left(-\frac{\partial^2 L}{\partial \theta \partial \theta^T} \right)^{-1} \frac{\partial L}{\partial \theta} \right]_{\theta^{(a)}} = \delta^{(a)} \quad (2.27)$$

El método de Newton falla ya que la matriz $-\frac{\partial^2 L}{\partial \theta \partial \theta^T}$ no es definida positiva para todo $\delta^{(a)}$. El algoritmo de Fisher reemplaza dicha matriz por su valor esperado

$$\mathbb{E} \left[-\frac{\partial^2 L}{\partial \theta \partial \theta^T} \right] = \mathbf{F}^T \mathbf{G} \mathbf{F} \quad (2.28)$$

donde

$$\mathbf{G} = \mathbb{E} \left[\frac{\partial L}{\partial f} \frac{\partial L}{\partial f^T} \right] \quad (2.29)$$

es una matriz definida positiva.

Esto da origen al algoritmo de Fisher

$$\theta^{(a+1)} = \theta^{(a)} + \left[(\mathbf{F}^T \mathbf{G} \mathbf{F})^{-1} \frac{\partial L}{\partial \theta} \right]_{\theta^{(a)}} \quad (2.30)$$

2.3 Incertidumbre, selección y promedio de modelos

En la sección previa, fueron presentados seis modelos no lineales los cuales pueden ser utilizados para la estimación de parámetros epidemiológicos y para hacer predicciones en tiempo real. En esta sección para tener en cuenta la incertidumbre de modelos, se describe el método de promedio de modelos (PM) [9, 10, 18], que combina las estimaciones de los diferentes modelos antes descritos. La idea es utilizar un promedio ponderado de las estimaciones de los parámetros obtenidas utilizando cada uno de los modelos alternativos en lugar de utilizar las estimaciones obtenidas con un solo modelo. En el marco del promedio de modelos, se ajusta un conjunto R modelos candidatos M_1, M_2, \dots, M_R al conjunto de datos para obtener las estimaciones de los parámetros de todos los modelos $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_R$. Un procedimiento post-selección [18, 51] implica que se necesita primeramente seleccionar el modelo con mejor bondad de ajuste a los datos, por ejemplo M_l , y estimar θ a través de $\hat{\theta}_l$. La selección del modelo puede estar basada en un criterio de información. Sin embargo, este procedimiento no tiene en

cuenta la incertidumbre de modelos porque la estimación está basada en un solo modelo. Las técnicas de promedio de modelos permiten estimar el componente en θ utilizando información obtenida de todos los modelos ajustados y en cierta forma tener en cuenta la incertidumbre de modelos. Supongamos que el Criterio de Información de Akaike (AIC , [1]) es utilizado para la selección de los modelos. Como los valores individuales AIC no son interpretables, ya que ellos contienen constantes arbitrarias y están muy afectados por el tamaño de la muestra, para un conjunto de R modelos candidatos M_1, M_2, \dots, M_R , Burnham y Anderson [9] propusieron rescalar los AIC a

$$\Delta AIC_i = AIC_i - AIC_{min}, \quad i = 1, \dots, R \quad (2.31)$$

Aquí, AIC_{min} es el valor de AIC más pequeño entre el conjunto de los R modelos. Las diferencias de AIC , ΔAIC_i , son interpretadas como la pérdida de información cuando el modelo M_i , en lugar del mejor modelo M_{min} , es utilizado para aproximar el modelo real desconocido. Burnham y Anderson [10] definieron los pesos de Akaike por:

$$w_i(AIC) = \frac{\exp(-\frac{1}{2}\Delta AIC_i)}{\sum_{i=1}^R \exp(-\frac{1}{2}\Delta AIC_i)}. \quad (2.32)$$

Los pesos de Akaike $w_i(AIC)$ pueden ser interpretados como los pesos de evidencia de que el modelo M_i es el mejor modelo dado un conjunto de R modelos y dado que uno de los modelos en el conjunto debe ser el mejor modelo. El modelo no lineal con el mayor peso de Akaike (es decir, el de mínimo AIC) es considerado como el modelo con mejor bondad de ajuste a los datos.

Basados en Burnham y Anderson [10], se puede calcular el estimador promedio de modelos para el acmé ($\hat{\eta}_{PM}$) y el tamaño final de la epidemia como sigue:

$$\hat{\eta}_{PM} = \sum_{i=1}^R w_i(AIC) \hat{\eta}_i, \quad (2.33)$$

$$\hat{\alpha}_{PM} = \sum_{i=1}^R w_i(AIC) \hat{\alpha}_i. \quad (2.34)$$

Donde $\hat{\eta}_i$ y $\hat{\alpha}_i$ son las estimaciones del acmé y del tamaño final de la epidemia del i -ésimo modelo respectivamente. El estimador para la varianza $\hat{\eta}_{PM}$ y $\hat{\alpha}_{PM}$ de los estimadores anteriores

están definidos respectivamente como:

$$\widehat{var}(\widehat{\eta}_{PM}) = \left[\sum_{i=1}^R w_i (AIC) \sqrt{\widehat{var}(\widehat{\eta}_i | M_i) + (\widehat{\eta}_i - \widehat{\eta}_{PM})^2} \right]^2, \quad (2.35)$$

$$\widehat{var}(\widehat{\alpha}_{PM}) = \left[\sum_{i=1}^R w_i (AIC) \sqrt{\widehat{var}(\widehat{\alpha}_i | M_i) + (\widehat{\alpha}_i - \widehat{\alpha}_{PM})^2} \right]^2. \quad (2.36)$$

Note que se puede reemplazar el criterio *AIC* por otro criterio de información tales como *BIC*, *KIC* y calcular los pesos de los modelos basados en estos criterios.

La deducción de estas las fórmulas se muestra para el estimador de la varianza $\widehat{\alpha}_{PM}$. Ignorando que los w_i necesitan ser estimado, la varianza de $\widehat{\alpha}_{PM}$ puede ser expresada como:

$$\begin{aligned} Var(\widehat{\alpha}_{PM}) = & \sum_{i=1}^R w_i^2 E \left[(\widehat{\alpha}_i - \alpha_{PM})^2 | M_i \right] \\ & + \sum_{i=1}^R \sum_{\substack{j=1 \\ i \neq j}}^R w_i w_j E \left[(\widehat{\alpha}_i - \alpha_{PM}) (\widehat{\alpha}_j - \alpha_{PM}) | M_i, M_j \right] \end{aligned} \quad (2.37)$$

El primer término, el error cuadrático medio de $\widehat{\alpha}_i$ dado el modelo M_i , puede ser escrito como:

$$E \left[(\widehat{\alpha}_i - \alpha_{PM})^2 | M_i \right] = E \left[\widehat{\alpha}_i - \alpha_i | M_i \right] + (\alpha_i - \alpha_{PM})^2 = Var(\widehat{\alpha}_i | M_i) + (\alpha_i - \alpha_{PM})^2 \quad (2.38)$$

Es decir, la suma de la varianza condicional $Var(\widehat{\alpha}_i | M_i)$ de $\widehat{\alpha}_i$ dado el modelo M_i y el sesgo al cuadrado de α_i con el parámetro promedio de modelo para el tamaño final de la epidemia α_{PM} . El término de la covarianza puede ser escrito como:

$$\begin{aligned} E \left[(\widehat{\alpha}_i - \alpha_{PM}) (\widehat{\alpha}_j - \alpha_{PM}) | M_i, M_j \right] &= \rho_{ij} \sqrt{E \left[(\widehat{\alpha}_i - \alpha_{PM})^2 | M_i \right] E \left[(\widehat{\alpha}_j - \alpha_{PM})^2 | M_j \right]} \\ &= \rho_{ij} \sqrt{\left(Var(\widehat{\alpha}_i | M_i) + (\alpha_i - \alpha_{PM})^2 \right) \left(Var(\widehat{\alpha}_j | M_j) + (\alpha_j - \alpha_{PM})^2 \right)} \end{aligned} \quad (2.39)$$

donde ρ_{ij} representa la correlación a través de modelo de $\widehat{\alpha}_i$ con $\widehat{\alpha}_j$, con respecto a α_{PM} . Por lo tanto,

$$\begin{aligned} Var(\widehat{\alpha}_{PM}) = & \sum_{i=1}^R w_i^2 \left[Var(\widehat{\alpha}_i | M_i) + (\alpha_i - \alpha_{PM})^2 \right] \\ & + \sum_{\substack{i=1 \\ i \neq j}}^R \sum_{j=1}^R w_i w_j \rho_{ij} \sqrt{\left(Var(\widehat{\alpha}_i | M_i) + (\alpha_i - \alpha_{PM})^2 \right) \left(Var(\widehat{\alpha}_j | M_j) + (\alpha_j - \alpha_{PM})^2 \right)} \end{aligned} \quad (2.40)$$

Si se asume que la correlación entre dos modelos es constante $\rho_{ij} = \rho$ y haciendo una pequeña transformación la formula se transforma en:

$$\begin{aligned} Var(\hat{\alpha}_{PM}) = & (1 - \rho) \left[\sum_{i=1}^R w_i^2 \left[Var(\hat{\alpha}_i | M_i) + (\alpha_i - \alpha_{PM})^2 \right] \right] \\ & + \rho \left[\sum_{i=1}^R w_i \sqrt{Var(\hat{\alpha}_i | M_i) + (\alpha_i - \alpha_{PM})^2} \right]^2 \end{aligned} \quad (2.41)$$

Si se asume correlación perfecta, es decir, $\rho = 1$, entonces se obtiene:

$$Var(\hat{\alpha}_{PM}) = \left[\sum_{i=1}^R w_i \sqrt{Var(\hat{\alpha}_i | M_i) + (\alpha_i - \alpha_{PM})^2} \right]^2 \quad (2.42)$$

Como los pesos de los modelos están basados en el *AIC* (o en cualquier otro criterio de información) el procedimiento de promedio de modelos descrito anteriormente asegura que las estimaciones de los parámetros acmé y tamaño final de la epidemia, y sus errores estándar, serán dominados por el modelo o los modelos con mejor bondad de ajuste.

2.3.1 Cuantificación de la incertidumbre de los parámetros

El procedimiento que se ha descrito puede ser utilizado para ajustar los modelos a la incidencia acumulada más que a la incidencia de intervalo correspondiente. Matemáticamente, las curvas acumuladas y de intervalo cargan la misma información, pero estadísticamente hablando, se esperan resultados diferentes de los dos enfoques. Los cálculos de la verosimilitud asume que los errores en las observaciones individuales son estadísticamente independientes. Este supuesto es particularmente inapropiado para curvas acumuladas, donde cada observación contiene todos los casos de las informaciones previas.

Para cuantificar la incertidumbre en la estimación de los parámetros fueron construidos intervalos de confianza generando conjuntos de realizaciones de la curva estimada por el promedio de modelos $C(t)$ utilizando bootstrap paramétrico [24]. Cada realización de número acumulado de casos notificados $C_i(t)$ ($i = 1, 2, \dots, m$) se genera como sigue: Para cada observación $C(t)$ para $t = 2, 3, \dots, n$ unidad de tiempo (por ejemplo, semana) se genera una nueva observación $C'_i(t)$ para $t \geq 2$ ($C'_1(t) = C_1(t)$) la cual es extraída de una muestra de la distribución Poisson con media $C(t) - C(t-1)$ (el incremento semanal en $C(t)$ desde la semana $t-1$ a la semana t). La realización correspondiente del número acumulado de notificaciones de dengue

está dado por $C_i(t) = \sum_{j=1}^t C'_i(j)$, donde $t = 1, 2, \dots, n$. De esta forma, el acmé y el tamaño final de la epidemia son estimados de cada una de las 1000 curvas epidémicas simuladas. La distribución del acmé y tamaño de la epidemia estimados se pueden utilizar para construir los intervalos de confianza del 95%.

2.3.2 Propuesta de procedimiento

En este epígrafe se propone un procedimiento para la estimación y predicción en tiempo real de los parámetros epidemiológicos, acmé y tamaño final de la epidemia. Se utilizaron varios modelos en el que los parámetros se estiman a través del método de promedio de modelos para tener en cuenta la incertidumbre de la selección del modelo en la estimación y en la inferencia. Este aspecto confirma la novedad de la presente investigación.

El procedimiento consiste en:

- Ajustar cada uno de los seis modelos no lineales tratados (Richards, logístico 3P, logístico 5P, Sigmoido Emax, Gompertz 4P y Weibull) al número acumulados de casos reportados en semanas sucesivas asumiendo distribución Normal asintótica y de Poisson. Para brotes de una sola onda, la estimación de los parámetros convergerá en cuanto la trayectoria se aproxime a la capacidad de carga.
- Tener un cuenta un criterio de información (AIC, BIC, KIC , etc.) para calcular los pesos de los modelos basados en estos criterios de acuerdo a las fórmulas 2.31 y 2.32 .
- Estimar o hacer predicciones en tiempo real puntuales y por intervalo de confianza de los parámetros, acmé y tamaño final de la epidemia, utilizando un promedio ponderado de modelos de acuerdo a las fórmulas 2.33-2.36.
- Hacer una validación de los escenarios considerados (análisis de estadísticos de bondad de ajuste, un análisis de los residuos, etc.)
- Construir intervalos de confianza para cuantificar la incertidumbre en la estimación de los parámetros generando conjuntos de realizaciones de la curva estimada por el

promedio de modelos utilizando bootstrap paramétrico, de acuerdo a los pasos explicado en el epígrafe 2.3.1.

Este procedimiento se implementó en el paquete de R, *DengueRT*, elaborado como parte de esta investigación y que se explica con detalles en el Capítulo 5.

2.4 Aplicación de la metodología

2.4.1 Estimación usando métodos del promedio de modelos

Todos los modelos discutidos anteriormente fueron ajustados al brote de dengue de una onda en las áreas "26 de Julio" y "J. R. Ramírez". Los modelos fueron ajustados al número acumulado de casos reportados semanales y los parámetros, acmé y tamaño final de la epidemia, fueron estimados para cada área utilizando el paquete de R, *DengueRT*. Las tablas 2.2 y 2.3 muestran las estimaciones de los parámetros, los criterios de información de Akaike y los pesos de Akaike para todos los modelos no lineales utilizados para calcular las estimaciones por promedio de modelos para el acmé y tamaño final de la epidemia, asumiendo que el número acumulado de casos reportados tienen distribución Normal y de Poisson respectivamente. Note que se obtuvo mejor bondad de ajuste cuando se asume que el número acumulado de casos reportados se distribuyen normalmente. Los otros resultados serán analizados considerando esta suposición. Las figuras 5.8 y 5.9 (Anexo 3) visualizan los valores predichos acumulados y de incidencia obtenido para los modelos ajustados a los datos completos en cada área de salud, junto con los valores observados.

Para el área "26 de Julio", el modelo de Richards, el modelo logístico 5P y el modelo logístico 3P tuvieron *AIC* notablemente más bajos que los modelos Sigmoido Emax, Gompertz y Weibull (Ver tabla 2.2). La estimación por promedio de modelos del tamaño final de este brote fue de 375.09 casos (IC 95%: 372.13, 378.05) con un acmé cerca de las 15.79 semanas (IC 95%: 15.43, 16.14) después del inicio del brote. Las estimaciones por intervalos de confianza para estos parámetros utilizando bootstrap fueron respectivamente (IC 95%: 373.61, 376.02) y (IC 95%: 15.63, 15.68). El modelo de Richards tiene un peso de Akaike igual a 0.9738, mientras

que los pesos obtenidos para los otros modelos son relativamente pequeños, lo que implica que la estimación por promedio de modelos estará dominada por el modelo de Richards para esta área de salud.

Tabla 2.2: Estimaciones de parámetros de interés obtenidas para los seis modelos no lineales y sus estimaciones por promedio de modelos asumiendo distribución Normal.

Áreas	Modelos	Acmé (η) (IC 95%)	Tamaño final(α) (IC 95%)	AIC	Pesos $w_i(AIC)$
26 de Julio	Richards	15.80 (15.55, 16.05)	375.08 (372.75, 377.42)	214.06	0.9738
	Logístico 3P	14.93 (14.79, 15.08)	378.29 (374.47, 382.10)	247.82	4.558E-08
	Logístico 5P	15.15 (15.03,15.27)	375.37 (372.13, 378.01)	221.29	0.0262
	Sigmoideo Emax	15.04 (14.85, 15.23)	380.35 (375.40, 385.29)	261.18	5.702E-11
	Gompertz 4P	13.98 (13.73, 14.23)	380.91 (374.49, 387.33)	280.71	3.277E-15
	Weibull	14.05 (13.77, 14.33)	385.42 (376.23, 394.61)	294.75	2.928E-18
	PM (IC 95%)	15.79 (15.43, 16.14)	375.09 (372.13, 378.05)	-	-
J. R. Ramirez	Richards	13.75 (13.29,14.20)	343.83 (339.90,347.76)	234.73	6.544E-04
	Logístico 3P	13.17 (12.99,13.35)	346.07 (341.79,350.35)	240.67	3.349E-05
	Logístico 5P	13.38 (13.23,13.53)	346.27 (342.63,349.92)	220.79	0.6971
	Sigmoideo Emax	13.36 (13.21,13.50)	348.00 (344.43,351.58)	222.46	0.3021
	Gompertz 4P	12.33 (12.14,12.52)	347.97 (343.60,352.35)	238.48	1.001E-04
	Weibull	12.37 (12.16,12.58)	353.66 (347.23,360.08)	250.10	3.003E-07
	PM (IC 95%)	13.37 (13.18, 13.56)	346.80 (341.81, 351.78)	-	-

Ahora se presentan los resultados para el área "J. R. Ramírez". En contraste con lo obtenido para el área "26 de Julio", el modelo logístico 5P fue encontrado ser el modelo con menor AIC ($AIC = 220.79$) y los valores de AIC de otros modelos no lineales están relativamente cerca. Se puede notar que los pesos de Akaike para los modelos logístico 5P, Sigmoideo Emax y Richards fueron 0.6971, 0.3021 y 6.544E-04 respectivamente. Por lo que las estimaciones de estos parámetros obtenidas por promedio de modelos son dominadas por la estimaciones obtenidas por estos tres modelos. El tamaño final observado en el área "J. R. Ramírez" es igual a 355 casos reportados. La estimación por promedio de modelos fue 346.80 casos (IC 95%:341.81,351.78) y la estimación para el acmé fue de 13.37 semanas (IC 95%:13.18,13.56). Las estimaciones por intervalos de confianza para estos parámetros utilizando bootstrap fueron respectivamente (IC 95%: 345.30, 347.52) y (IC 95%: 13.37, 13.40).

Tabla 2.3: Estimaciones de parámetros de interés obtenidas para los seis modelos no lineales y sus estimaciones por promedio de modelos asumiendo distribución Poisson.

Áreas	Modelos	Acme (η) (IC 95%)	Tamaño final(α) (IC 95%)	AIC	Pesos $w_i(AIC)$
26 de Julio	Richards	15.85 (15.30, 16.40)	375.01 (364.75, 385.27)	248.6	0.9987
	Logístico 3P	15.08 (14.76, 15.40)	383.57 (373.26, 393.88)	262.0	0.0012
	Logístico 5P	15.08 (14.77, 15.38)	377.10 (367.05, 387.15)	269.4	3.039E-05
	Sigmoideo Emax	15.05 (14.70, 15.40)	389.99 (377.85, 402.13)	288.6	2.058E-09
	Gompertz 4P	13.68 (13.38, 13.98)	392.32 (380.09, 404.56)	310.9	2.959E-14
	Weibull	16.27 (15.90, 16.65)	377.10 (367.05, 387.15)	267.4	8.261E-05
	PM (IC 95%)	15.85 (15.16, 16.54)	375.02 (361.99, 388.04)	-	-
J. R. Ramirez	Richards	14.51 (13.92, 15.10)	341.74 (331.88, 351.60)	251.6	0.92572
	Logístico 3P	13.45 (13.09, 13.81)	354.01 (343.68, 364.35)	277.8	1.893E-06
	Logístico 5P	13.43 (13.11, 13.76)	343.29 (333.17, 353.41)	259.6	0.022
	Sigmoideo Emax	13.39 (13.04, 13.74)	351.97 (340.56, 363.37)	262.7	0.004
	Gompertz 4P	12.23 (11.92, 12.53)	352.91 (341.66, 364.16)	274.0	1.266E-05
	Weibull	14.50 (14.10, 14.90)	342.94 (333.27, 352.62)	257.5	0.048
	PM (IC 95%)	14.48 (13.69, 15.27)	341.87 (329.31, 354.43)	-	-

2.4.2 Predicción en tiempo real

La metodología propuesta por promedio de modelos es particularmente útil para la predicción en tiempo real ya que se utiliza solo parte de los datos para predicciones a largo plazo. En este caso, no se quiere basar la estimación en un solo modelo. Las estimaciones puntuales y por intervalos de confianza al 95% para el acmé y tamaño final de la epidemia para todos los modelos no lineales y por promedio de modelos usando diferentes intervalos de tiempo en las áreas de salud "26 de Julio" y "J. R. Ramirez" se presentan en las tablas 5.2, 5.3, 5.4 y 5.5 (Anexo 3) respectivamente.

Para el área "26 de Julio" la estimación por promedio de modelos para el acmé de la epidemia se estabiliza a partir de la semana 19 en adelante (ver tabla 5.2). Esto indica que las autoridades de salud pudieran detectar el acmé de la epidemia 3 semanas después de su ocurrencia. Además, las estimaciones por promedio de modelos para el tamaño final de la epidemia se estabiliza alrededor de las semanas 19 ó 20 (ver tabla 5.3). Esto indica que aproximadamente 4 meses antes del fin de la epidemia esta valiosa información puede estar disponible para las

autoridades de salud.

Para el área de "J. R. Ramírez", la estimación por promedio de modelos para el acmé se estabiliza en la semana 16-17 (ver la tabla 5.4 y el panel de la derecha de la figura 2.1). La estimación puntual para el acmé en la semana 17, 13.36, es muy similar a la estimación si son utilizados todos los datos 13.37 (ver la tabla 5.4). El tamaño final observado de la epidemia es 355 casos. Los modelos logístico 5P, Sigmoideo Emax y el modelo de Richards (con pesos de Akaike de 0.6971, 0.3021 y 6.544E-04, respectivamente) todos subestiman el tamaño final de la epidemia. Como resultado, la estimación por promedio de modelos se estabiliza lentamente a la estimación 346.88 (IC 95%:341.81,351.78) de la semana 34 (ver la tabla 5.5, el panel de la izquierda de las figuras 2.1 y 2.2). Note que si se utiliza el límite superior del intervalo de confianza para el tamaño final como peor escenario, la estimación por promedio de modelos para el límite superior está en el rango 365.51 en la semana 17 hasta 351.78 en la semana 34. Por lo tanto, en la mitad de la epidemia pudieran ser reportadas a las autoridades de salud, estimaciones precisas del tamaño final de la epidemia.

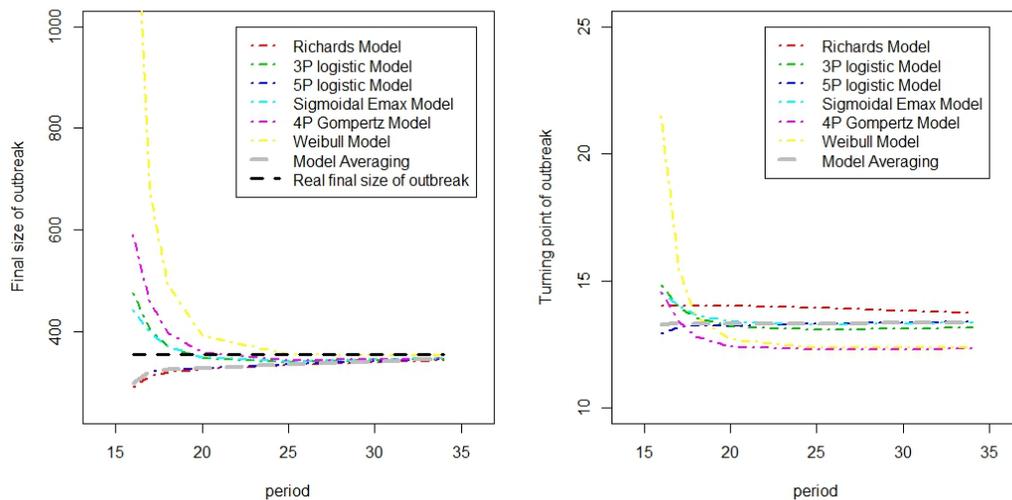


Figura 2.1: Cambios en el tiempo de las estimaciones del tamaño final y acmé de la epidemia utilizando todos los modelos y por promedio de modelos en el área "J. R. Ramírez".

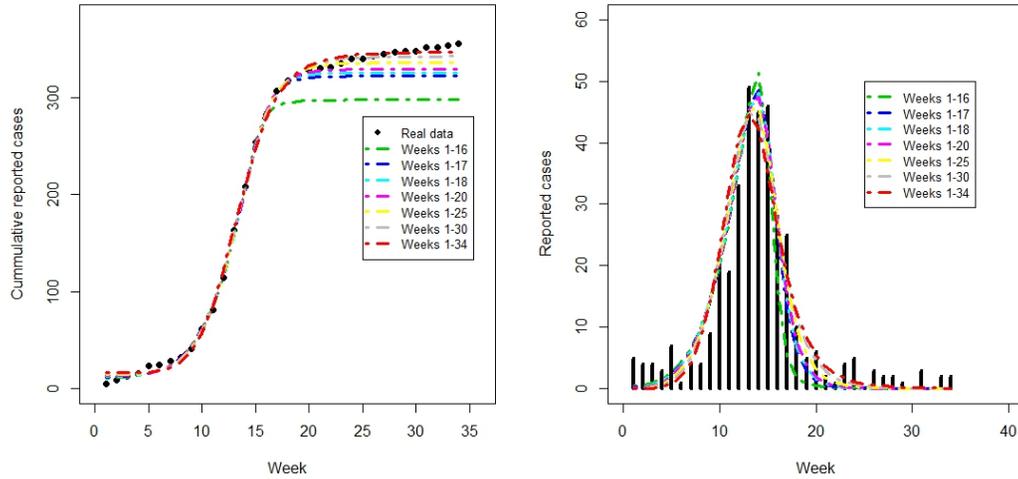


Figura 2.2: Predicciones por promedio de modelos para las curvas epiémicas acumulativa y de incidencia usando diferentes períodos de tiempo en el área "J. R. Ramírez".

Nota Durante la realización de esta investigación se realizó una generalización de la metodología propuesta a brotes de Zika en varias ciudades de Colombia, cuyos resultados fueron publicado en la Revista *Epidemiology & Infection*. Algunas salidas de la aplicación de la metodología propuesta a estos datos de brotes de Zika en Colombia son mostrados en las figuras 5.10 y 5.11 (Anexo 4).

2.5 Conclusiones parciales

La modelación de un brote de una onda requiere usar un modelo no lineal de crecimiento para estimar los parámetros epidemiológicos de interés. En este estudio hemos mostrado que varios modelos no lineales; Richards, logístico 3P, logístico 5P, Sigmoideo Emax, Gompertz 4P y Weibull; pueden ser utilizados para modelar los datos. Alternativamente, se puede aplicar una metodología de modelos promedios que use una estimación de parámetros ponderadas basada en la probabilidad posterior. Es esta investigación, se aboga por el uso de la metodología de promedio de modelos ya que no ignora la incertidumbre relacionada a la selección de modelos, la cual es ignorada cuando se conduce una inferencia y estimación post selección.

Se ha mostrado que la metodología de promedio de modelos puede ser utilizada para llevar a cabo estimaciones en tiempo real del acmé y predicciones para el tamaño final de la epidemia. En la aplicación al brote del 2001/2002, estimaciones relativamente precisas para estos parámetros pudieran ser dadas a las autoridades de salud cuatro meses antes del fin de la epidemia. Además, para las predicciones en tiempo real la metodología de promedio de modelo ofrece un atractivo enfoque de modelación ya que los datos disponibles representan solo una parte de los datos del brote y tiene en cuenta un conjunto de modelos y las predicciones en tiempo real son dominadas por el modelo (los modelos) con mejor bondad de ajuste a los datos. Esa es la principal ventaja del uso de la metodología de promedio de modelos. Por eso, según nuestra opinión, es apropiado en esta situación tener en cuenta varios modelos posibles.

Predecir la tendencia de una epidemia a partir de datos limitados durante las etapas tempranas de una epidemia puede ser algunas veces engañoso. No obstante, la predicción temprana de la magnitud de un brote epidémico es más importante que los estudios retrospectivos. La metodología propuesta en este capítulo no permite la predicción del acmé de la epidemia sino la estimación. Esto es debido al hecho que los modelos no lineales presentados en este capítulo ajustan pobremente los datos en las etapas iniciales de los brotes, es decir, cuando el acmé puede ser predicho. Como se observó en otros momentos en la predicción en tiempo real, el pronóstico parece ser muy vulnerable al tiempo de predicción, especialmente durante las fases tempranas de la epidemia. Tan pronto los datos incluyen este punto de inflexión y un corto intervalo de tiempo después, el ajuste de la curva y las predicciones futuras del número de casos serán bastante precisas. A pesar de esta limitación, el presente estudio ofrece substancialmente una estrategia de modelación para implementar pronósticos en tiempo real de una epidemia en el medio de su curso.

Capítulo 3

Modelación de brotes de una onda en diferentes regiones

3.1 Introducción

El capítulo anterior se centró en la modelación de brotes de dengue una onda basado en casos acumulados en una región determinada con el objetivo de estimar parámetros epidemiológicos claves. En este capítulo se proponen diferentes métodos estadísticos para la estimación de parámetros epidemiológicos en brotes de una onda pero cuando están coleccionados datos de casos acumulados para dos o más regiones.

Un aspecto que pudiera interesar a los epidemiólogos es la comparación de brotes de dengue basado en las estimaciones de parámetros epidemiológicos cuando se cuentan con los datos de casos acumulados de dengue en dos regiones o con los datos de la misma región en dos momentos determinados. Este es un tema que no ha sido abordado con profundidad en la literatura relacionada con brotes de dengue. En este capítulo se proponen diferentes enfoques metodológicos basados en abordajes Frecuentista y Bayesiano que permitan tener en cuenta la incertidumbre de modelos en la comparación de brotes epidémicos de dengue, basado en la estimación de los tres parámetros epidemiológicos claves: el acmé de la epidemia, el tamaño final y el número reproductivo básico R_0 . Los métodos Frecuentistas hacen una extensión de los modelos logístico de tres parámetro y el de Richards y utilizan el promedio de modelos. Los métodos Bayesianos utilizan el modelo Bayesiano jerárquico y además, se utiliza un promedio

Bayesiano de modelos aplicando selección Gibbs de variable.

Algunas veces los reportes de casos acumulados se coleccionan para todas las áreas urbanas en una región en particular, y para los epidemiólogos resulta de interés estimar los parámetros epidemiológicos claves o estudiar la heterogeneidad regional de la transmisión de la enfermedad. La modelación, para estimar parámetros epidemiológicos, se realiza usualmente para cada área por separado, pero cuando el interés consiste en estimar el comportamiento promedio de un área en particular en la población, y la variabilidad entre y dentro de las áreas, se propone un modelo no lineal con efectos mixtos. En este capítulo, se describen dos enfoques que proporcionan estimaciones de tres parámetros epidemiológicos claves. El primer enfoque, el clásico, consiste en ajustar un modelo no lineal individual para cada área por separado. En el segundo enfoque, la propuesta, se utiliza un modelo no lineal con efectos mixtos, el cual tiene en cuenta la heterogeneidad entre las áreas.

3.2 Incertidumbre de modelos en la comparación de dos brotes

3.2.1 Enfoque Frecuentista

Sea Y_t una respuesta aleatoria que representa el número acumulado de casos reportados en el tiempo t . En este caso, se considera el supuesto que la respuesta Y_t tiene distribución normal asintótica con media $\mu(t, \theta)$ y varianza σ^2 , es decir, $Y_t \sim N(\mu(t, \theta), \sigma^2)$.

Como en el capítulo anterior, para describir la relación entre el número acumulado de casos reportados y el tiempo y para estimar parámetros epidemiológicos claves se pueden utilizar varios modelos no lineales. En este caso se supone que la estructura de media es el modelo logístico 3P o el modelo de Richards (Ver Tabla 2.1) porque con estos modelos es posible estimar el parámetro epidemiológico más importante R_0 .

Como el objetivo es comparar los tres parámetros epidemiológicos de los dos brotes de dengue, estos modelos fenomenológicos no distinguen entre los brotes. En esta sección se extienden estos modelos y se propone un modelo que tenga en cuenta ambos brotes. Para incorporar los

dos brotes en el modelo, se define una variable dicotómica que toma el valor 1 para un brote y cero para el otro, esto es

$$I_t = \begin{cases} 0 & \text{1. brote} \\ 1 & \text{2. brote} \end{cases} \quad (3.1)$$

Además, se asume una dependencia lineal de los parámetros del modelo. En el caso donde se considera como estructura de la media al modelo logístico de tres parámetros, la dependencia lineal de los modelos del parámetro puede ser expresada de la siguiente forma:

$$\begin{aligned} \alpha &= \alpha_1 + \beta_1 * I_t \\ \gamma &= \gamma_1 + \beta_2 * I_t \\ \eta &= \eta_1 + \beta_3 * I_t \end{aligned} \quad (3.2)$$

Note que los parámetros β_1, β_2 y β_3 pueden ser interpretados como las diferencias entre el tamaño final, la razón de crecimiento y el acmé de la epidemia entre los dos brotes respectivamente.

Similarmente, cuando se considere como estructura de la media al modelo de Richards, la dependencia lineal de los parámetros del modelo puede expresarse como:

$$\begin{aligned} \alpha &= \alpha_1 + \beta_1 * I_t \\ \gamma &= \gamma_1 + \beta_2 * I_t \\ \eta &= \eta_1 + \beta_3 * I_t \\ k &= k_1 + \beta_4 * I_t \end{aligned} \quad (3.3)$$

Una ventaja de este modelo, es que permite probar la diferencia entre los parámetros epidemiológicos claves de los dos brotes de dengue. Formalmente, para cada uno de los parámetros se puede probar la hipótesis:

$$\begin{aligned} H_{0,i} &: \beta_i = 0 \\ H_{1,i} &: \beta_i \neq 0 \end{aligned} \quad (3.4)$$

Para estimar los parámetros de este modelo se pueden utilizar los mismos métodos descritos en el capítulo 2. Para estimar estos parámetros y probar estas hipótesis se pueden auxiliar del procedimiento NLMIXED del software SAS.

Promedio de modelos: enfoque tradicional

Al modelo expuesto previamente (M_1) se le pueden imponer ciertas restricciones dependiendo de las diferentes combinaciones con la presencia o no de los β_i . Cuando se considera como estructura de la media el modelo logístico de tres parámetros, se obtienen 8 modelos M_k , $k = 1, 2, \dots, 8$ (Ver Tabla 3.2.1). En el caso, donde se considere el modelo de Richards como estructura de media se obtienen 16 modelos.

Modelos (M_k)	Restricciones
M_1	-
M_2	$\beta_2 = 0$
M_3	$\beta_3 = 0$
M_4	$\beta_2 = 0, \beta_3 = 0$
M_5	$\beta_1 = 0$
M_6	$\beta_1 = 0, \beta_2 = 0$
M_7	$\beta_1 = 0, \beta_3 = 0$
M_8	$\beta_1 = 0, \beta_2 = 0, \beta_3 = 0$

Con estos modelos se puede llevar a cabo el procedimiento de promedio de modelos expuesto en el capítulo 2, por lo que esta variante asegura que las estimaciones de los parámetros sean dominados por el modelo o los modelos con mejor bondad de ajuste.

3.2.2 Enfoque Bayesiano

Hasta aquí los parámetros de los modelos para la comparación de los brotes han sido estimado dentro del marco Frecuentista. En este epígrafe se propone estimar los parámetros dentro del marco Bayesiano.

Modelo jerárquico Bayesiano

Los modelos M_k se pueden ajustar y estimar los parámetros basados en un enfoque Bayesiano, para ello se propone utilizar un modelo Bayesiano jerárquico. Sea Y_t el número acumulado de casos reportados en el tiempo t una variable con distribución normal asintótica. Por lo

que se asume el siguiente modelo jerárquico normal para esta variable. En la primera etapa del modelo se asume una verosimilitud Normal (se considera a modo de ilustración cuando la estructura de media es el modelo logístico 3P),

$$Y_t \sim N(\mu(t), \tau) \quad (3.5)$$

$$\mu[t] = \frac{\alpha + \beta_1 * I[t]}{1 + e^{-(\gamma + \beta_2 * I[t]) * (t - \eta - \beta_3 * I[t])}}$$

En la segunda etapa del modelo se especifican distribuciones *a priori* no informativas independientes para los parámetros. La segunda etapa puede ser expresada como:

$$\begin{aligned} \alpha &\sim N(0, \tau_\alpha) \\ \gamma &\sim N(0, \tau_\gamma) \\ \eta &\sim N(0, \tau_\eta) \end{aligned} \quad (3.6)$$

Para completar las especificaciones del modelo de probabilidad, en la tercera etapa se asume la distribución de los hiperparámetros:

$$\tau, \tau_\alpha, \tau_\gamma, \tau_\eta \sim \text{gamma}(0.001, 0.001) \quad (3.7)$$

Utilizando el modelo Bayesiano jerárquico, se puede tener una segunda variante para comparar los parámetros de ambos brotes teniendo en cuenta la incertidumbre de modelos. Esta variante consistiría en aplicar la técnica de promedio de modelos a los 8 modelos M_k , $k = 1, 2, \dots, 8$ (ó 16 si se utiliza el modelo de Richards) utilizando el criterio de información de desviación (DIC) para el cálculo de los pesos según las fórmulas 2.31 y 2.32 presentadas en el capítulo 2.

Una vez especificado el modelo Bayesiano jerárquico, este puede ser ajustado usando los métodos de Monte Carlo con Cadenas de Markov (MCMC). Los métodos MCMC proporcionan un desarrollo unificado dentro del cual muchos problemas complejos pueden ser analizados usando un software genérico [30]. Existen muchas formas de construir cadenas de Markov, pero los métodos más usados son: el algoritmo Metropolis-Hastings y el algoritmo muestreador de Gibbs. Estos algoritmos se encuentran implementados en el software WinBUGS .

Algoritmos Metropolis-Hastings

El algoritmo de Metropolis-Hastings es un método MCMC para obtener una sucesión de muestras aleatorias de una función de densidad de probabilidad de la cual, el muestreo directo

se dificulta. Esta sucesión se constituye para aproximar la distribución *a posteriori* en el caso de la inferencia Bayesiana . La teoría de este algoritmo puede consultarse en Casella y Robert [12] o en Liang y cols. [49].

El algoritmo consiste en generar valores de una cadena de Markov cuya distribución estacionaria, distribución marginal de la cadena de Markov, sea la distribución objetivo $\pi(\theta)$. Esto permite buscar entre ellas, una que además posea otras propiedades complementarias como, rapidez de convergencia a estacionariedad y no demasiada autocorrelación.

El algoritmo construye una cadena de Markov apropiada definiendo las probabilidades de transición de la siguiente forma. Sea $Q(\theta^*|\theta)$ una distribución de transición (arbitraria) y sea

$$\alpha(\theta^*, \theta) = \min \left\{ \frac{\pi(\theta^*|y)Q(\theta|\theta^*)}{\pi(\theta|y)Q(\theta^*|\theta)}, 1 \right\}$$

Algoritmo: Dado un valor inicial $\theta^{(0)}$, la t -ésima iteración consiste en:

1. generar una observación θ^* de $Q(\theta^*|\theta^{(t)})$,
2. generar una variable $u \sim U(0, 1)$;
3. si $u \leq \alpha(\theta^*, \theta)$, hacer $\theta^{(t+1)} = \theta^*$; en caso contrario, hacer $\theta^{(t+1)} = \theta^{(t)}$.

Este procedimiento genera una cadena de Markov con distribución de transición:

$$\pi(\theta^{(t+1)}|\theta^{(t)}) = \alpha(\theta^{(t+1)}, \theta^{(t)})Q(\theta^{(t+1)}|\theta^{(t)})$$

La probabilidad de aceptación $\alpha(\theta^*, \theta)$, solo depende de $\pi(\theta|y)$ a través de un cociente, por lo que la constante de normalización no es necesario conocerla. Por ello, el algoritmo de Metropolis-Hastings puede ser una herramienta muy útil en la aplicación de las técnicas Bayesianas, donde la constante de integración de la distribución *a posteriori* no suele ser conocida.

Muestreador de Gibbs

Entre las técnicas MCMC el algoritmo de Gibbs es uno de los métodos más fáciles de aplicar y sin duda, el más conocido y utilizado. En Casella y George [11] se da una introducción clara y concisa de este método, y en Gelfand y cols. [28] se presentan aplicaciones del mismo.

El muestreador Gibbs aproxima integrales que no pueden calcularse en forma analítica, generando cadenas de Markov Monte Carlo (MCMC), donde la transición del origen de la distribución $\pi(\theta)$ está formada por las distribuciones condicionales completas ($p(\theta_i) = p_i(\theta_i|\theta_{-i})$). Se asume que la distribución de interés es $\pi(\theta)$, donde el vector θ puede descomponerse en $k \geq 2$ subvectores, $\theta = (\theta_1, \dots, \theta_k)$. Cada uno de los componentes θ_i de θ puede ser un escalar, un vector o una matriz. Se considera que las distribuciones condicionales completas $\pi_1(\theta_i) = \pi_1(\theta_i|\theta_{-i})$ están disponibles, pudiendo generar valores de las mismas sin excesivo costo computacional, siendo θ_i el vector $(\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$.

El objetivo del algoritmo de Gibbs es generar una muestra suficientemente grande para ajustar una adecuada distribución *a posteriori*. A partir de ella se podrán hacer inferencias sobre los momentos, las marginales, la distribución predictiva, o cualquier otra característica de la distribución posterior que sea de interés. El problema que se debe resolver es cómo tomar una muestra de la distribución π , cuando los planes para la generación de las muestras son complicados o simplemente no se conoce el origen de la distribución π , pero es posible generar muestras de las distribuciones $\pi_i(\theta_i)$.

El algoritmo Gibbs puede ser descrito de la siguiente forma:

1. Se inicia el contador de la iteración de la cadena en $j = 1$ y se asignan valores iniciales para el vector $\theta^{(0)} \leftarrow (\theta_1^{(0)}, \dots, \theta_k^{(0)})$

2. Repetir hasta lograr la convergencia

{

Se obtiene un nuevo valor $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_k^{(j)})'$ a partir de $\theta^{(j-1)}$ por la sucesiva generación de los valores:

$$\theta_1^{(j)} \sim \pi_1(\theta_1|\theta_2^{(j-1)}, \dots, \theta_k^{(j-1)})$$

$$\theta_2^{(j)} \sim \pi_k(\theta_2 | \theta_1^{(j)} \theta_3^{(j-1)}, \dots, \theta_k^{(j-1)})$$

⋮

$$\theta_k^{(j)} \sim \pi_k(\theta_k | \theta_1^{(j)}, \dots, \theta_{k-1}^{(j)})$$

3. Se actualiza el contador de j a $j + 1$ y se regresa al paso 2.

}

Cuando la cadena converge, los valores resultantes de $\theta^{(j)}$ son una muestra de la distribución π . Se asume la convergencia de la cadena si misma se aproxima a una condición de equilibrio cuando el número de iteraciones se incrementa [47, 53]. Las condiciones de convergencia para el muestreador de Gibbs fueron establecidas por Smith y Robert [80].

Un valor de la distribución de interés π se obtiene solamente cuando el número de iteraciones de la cadena se aproxima a infinito. En la práctica esto no es posible y un valor obtenido en una iteración suficientemente grande se toma como una muestra de la distribución π . La dificultad es la determinación de cuán grande deberá ser el número de iteraciones. No hay respuesta simple a esta pregunta y los mayores esfuerzos se han orientado al estudio de las características de la convergencia de las cadenas [29].

La muestra obtenida de θ (con $j \rightarrow \infty$) es una muestra correlacionada de la distribución *a posteriori* de la cual, se puede obtener cualquier cantidad de interés. Usando la estimación Monte Carlo es posible encontrar:

$$\widehat{E}(\theta_i | y) = \frac{1}{K - j_0} \sum_{j=j_0+1}^K \theta_i^{(j)}$$

La iteración de $j = 0$ a $j = j_0$ se conoce como período de inicialización (*burn-in*). En la práctica, se pueden simular paralelamente m cadenas, en este caso, el estimador *a posteriori* de la media será:

$$\widehat{E}(\theta_i | y) = m \frac{1}{K - j_0} \sum_{i=1}^m \sum_{j=j_0+1}^K \theta_{i,j}^{(j)}$$

Promedio de modelo Bayesiano: Selección Gibbs de variables

Una tercera variante, completamente Bayesiana, para la comparación de brotes teniendo en cuenta la incertidumbre de modelos consiste en conducir un promedio Bayesiano de modelos y comparar los resultados utilizando una técnicas de simulación: selección Gibbs de variable (GVS). Si se introduce la función indicadora variable $g(\cdot)$, donde $g(j) = 1$ si β_j está en el modelo y $g(j) = 0$ si β_j no está en el modelo, se reduce el marco a uno de dimensionalidad fija [19]. Se pueden utilizar técnicas de simulación para estimar $g(\cdot)$ y los parámetros para todos los modelos, M_k , $k = 1, 2, \dots, 8$. Esta propuesta puede ser implementada utilizando el software WinBUGS. A continuación se muestran dos casos para distribuciones *a priori* diferentes:

Verosimilitud

$$Y[t] \sim N(\mu[t], \tau) \quad (3.8)$$

$$\mu[t] = \frac{\alpha + g(1) * \beta_1 * I[t]}{1 + e^{-(\gamma + g(2) * \beta_2 * I[t]) * (t - \eta - g(3) * \beta_3 * I[t])}}$$

A priori (Caso 1)

$$g(j) \sim \text{Bernoulli}(0.5) \quad \text{for } j = 1, 2, 3$$

$$\alpha \sim N(0, \tau_1), \gamma \sim N(0, \tau_2), \eta \sim N(0, \tau_3) \quad (3.9)$$

$$\tau, \tau_1, \tau_2, \tau_3 \sim \text{gamma}(0.001, 0.001)$$

A priori (Caso 2)

$$g(j) \sim \text{Bernoulli}(\pi_j) \quad \text{for } j = 1, 2, 3$$

$$\pi_j \sim U(0, 1)$$

$$\alpha \sim N(0, \tau_1), \gamma \sim N(0, \tau_2), \eta \sim N(0, \tau_3) \quad (3.10)$$

$$\tau, \tau_1, \tau_2, \tau_3 \sim \text{gamma}(0.001, 0.001)$$

Note que esta ultima variante tiene la ventaja que permite estimar simultáneamente los pesos de todos los posibles modelos y los parámetros.

3.2.3 Aplicación: Comparación brotes dengue, La Lisa 2006 y 2007

En esta sección se aplican los métodos expuestos previamente para la comparación de los parámetros epidemiológicos claves de dos brotes de dengue ocurridos en el municipio La Lisa en el 2006 y 2007. En esta aplicación se considera que la estructura de media es el modelo

logístico 3P. Primeramente se utilizaron los métodos propuestos del enfoque Frecuentista, la cual consistía en utilizar los modelos M_k , $k = 1, 2, \dots, 8$ y aplicar el procedimiento de promedio de modelos expuestos en el capítulo 2. En la tabla 3.1 se muestran los criterios de información AIC y BIC para los 8 modelos así como los pesos o probabilidades *a posteriori*. También en la tabla 3.1 se encuentra el criterio de información de desviación y las probabilidades *a posteriori* cuando los modelos son ajustados utilizando el modelo Bayesiano jerárquico y aplicando el promedio de modelos con este criterio. Note como las estimaciones para cada uno de los criterios de información son dominadas por el modelo M_2 , es decir cuando $\beta_2 = 0$, por lo que las estimaciones de los parámetros serán muy similares a las obtenidas por este modelo.

Tabla 3.1: El AIC, BIC y DIC para la selección de modelos. Probabilidades estimadas *a posteriori* utilizando aproximaciones AIC, BIC and DIC.

Modelos (M_k)	Restricciones	AIC	BIC	DIC	$p_k(\text{AIC})$	$p_k(\text{BIC})$	$p_k(\text{DIC})$
M_1	-	256.9	268.3	257.646	0.347293	0.18932	0.322366
M_2	$\beta_2 = 0$	255.8	265.7	256.330	0.601946	0.694671	0.622464
M_3	$\beta_3 = 0$	263.3	273.1	263.742	0.014156	0.017175	0.015297
M_4	$\beta_2 = 0, \beta_3 = 0$	261.4	269.6	261.826	0.036604	0.098834	0.039872
M_5	$\beta_1 = 0$	336.1	345.9	352.179	2.2E-18	2.67E-18	9.57E-22
M_6	$\beta_1 = 0, \beta_2 = 0$	400.5	408.7	401.039	2.28E-32	6.16E-32	2.35E-32
M_7	$\beta_1 = 0, \beta_3 = 0$	458.4	466.6	462.527	6.1E-45	1.65E-44	1.04E-45
M_8	$\beta_1 = 0, \beta_2 = 0, \beta_3 = 0$	468.9	475.5	472.316	3.2E-47	1.92E-46	7.82E-48

La tabla 3.2 muestra las estimaciones de los parámetros para esta variante cuando se utilizan los pesos de Akaike (AIC). Note como el coeficiente β_2 no es significativo indicando que si se utiliza este modelo para la inferencia, la razón de crecimiento y por ende la transmisión sería la misma en ambos brotes.

Finalmente se muestran los resultados cuando se conduce un promedio Bayesiano de modelos utilizando la selección Gibbs de variable. La tabla 3.3 muestra las probabilidades a posteriores de los modelos con las distribuciones *a posteriores* según los casos 1 y 2. Note que nuevamente, las mayores probabilidades *a posteriori* corresponden con el modelo M_2 , aunque ligeramente menores que con los otros criterios.

Tabla 3.2: Estimaciones puntuales y por intervalo de confianza al 95% de los parámetros utilizando la variante de promedio de modelos con los pesos AIC.

Parámetros	Estimaciones (IC 95%)	Sig ($Pr > t $)
α_{06}	493.49 (487.17,499.81)	< .0001
α_{07}	186.07 (178.63,193.52)	< .0001
γ_{06}	0.5203 (0.4967,0.5439)	< .0001
γ_{07}	0.4871 (0.4249,0.5493)	< .0001
η_{06}	9.3067 (9.1963,9.4171)	< .0001
η_{07}	9.8263 (9.4869,10.1657)	< .0001
$R_{0,06}$	4.1056 (3.8427,4.3685)	< .0001
$R_{0,07}$	3.7517 (3.1183,4.3851)	< .0001
β_1	-307.42 (-317.18,-297.65)	< .0001
β_2	-0.0332 (-0.0997,0.0333)	0.3186
β_3	0.5196 (0.1627,0.8765)	0.0055
$R_{0,06} - R_{0,07}$	0.3539 (-0.3319,1.0397)	0.3028

La tabla 3.4 muestra las estimaciones de la media de las funciones indicadoras $g(j)$ para los casos 1 y 2, además de las medias de π en el caso 2 para cada uno de los j . Observe como para cada caso, la media de $g(1) = 1$ indicando que los tamaños finales de las epidemias son

Tabla 3.3: Probabilidades estimadas *a posteriori* con distribuciones *a priori* (Caso 1 y 2), utilizando selección Gibbs de variable.

Modelos (M_k)	Expresión	$p_k(\text{GVS})$ (Caso 1)	$p_k(\text{GVS})$ (Caso 2)
M_1	$g(1) * g(2) * g(3)$	0.1864744	0.159195
M_2	$g(1) * [1 - g(2)] * g(3)$	0.5639256	0.482205
M_3	$g(1) * g(2) * [1 - g(3)]$	0.0620256	0.089005
M_4	$g(1) * [1 - g(2)] * [1 - g(3)]$	0.1875744	0.269595
M_5	$[1 - g(1)] * g(2) * g(3)$	0	0
M_6	$[1 - g(1)] * [1 - g(2)] * g(3)$	0	0
M_7	$[1 - g(1)] * g(2) * [1 - g(3)]$	0	0
M_8	$[1 - g(1)] * [1 - g(2)] * [1 - g(3)]$	0	0

significativamente diferentes con total certeza.

Tabla 3.4: Las medias de $g(j)$ (caso 1 y caso 2) y las medias de π (caso 2) utilizando la selección Gibbs de variables.

$g(j)$	Media (Caso 1)	Media (Caso 2)	$\bar{\pi}$ (Caso 2)
$g(1)$	1	1	0.6657
$g(2)$	0.2485	0.2482	0.4147
$g(3)$	0.7504	0.6414	0.5469

Las estimaciones puntuales y por intervalo de confianza al 95% para cada uno de los parámetros epidemiológicos claves en los dos brotes y para cada caso utilizando la selección Gibbs de variable se muestran en la tabla 3.5. Note como de acuerdo a estas estimaciones los parámetros β_2 y β_3 no son significativos para cada uno de los casos, indicando que existen diferencias significativas entre el acmé y tamaño final de ambos brotes.

Tabla 3.5: Estimación e intervalos creíbles al 95% de los parámetros para los casos 1 y 2, utilizando selección Gibbs de variable.

Parámetros	Estimaciones (95% CI) (Caso 1)	Estimaciones (95% CI) (Caso 2)
α_{06}	494.6 (487.6,501.9)	494.4 (487.5,501.5)
α_{07}	182.4 (175.3,190.0)	182.6 (175.7,190.0)
η_{06}	9.332 (9.205,9.462)	9.328 (9.207,9.455)
η_{07}	9.474 (4.736,11.35)	9.627 (9.012,10.05)
$R_{0,06}$	4.074 (3.807, 4.366)	4.077 (3.813,4.365)
$R_{0,07}$	4.536 (1.361, 12.29)	4.526 (1.343,12.22)
β_1	-312.2 (-322.1,-302.0)	-311.8 (-321.4,-301.9)
β_2	-0.00195 (-0.4041,0.4077)	-0.00305 (-0.4084,0.405)
β_3	0.1422 (-4.639,1.989)	0.299 (-0.3405,0.7467)

3.3 Modelación de datos de varias áreas en una misma región

Hasta ahora hemos analizado los métodos de estimación de parámetros epidemiológicos cuando tenemos los datos de dos brotes de dengue. Sin embargo, en algunas veces los reportes de casos

acumulados se coleccionan para todas las áreas urbanas en una región en particular, y para los epidemiólogos resulta de interés estimar los parámetros epidemiológicos claves o estudiar la heterogeneidad regional de la transmisión de la enfermedad. En esta sección se describen los métodos de estimación: el enfoque clásico, la propuesta y se muestra la aplicación de ambos enfoques

3.3.1 Metodología clásica: Modelo para cada área específica

Sea Y_{ij} una respuesta aleatoria, que representa el número acumulado de casos reportados en el área de salud i en el tiempo t_j . En este estudio en un primer escenario se asume que el número acumulado de casos notificados, la respuesta (Y_{ij}), tiene distribución normal asintótica con media $\mu(t_j, \theta_i)$ y varianza σ^2 , $Y_{ij} \sim N(\mu(t_j, \theta_i), \sigma^2)$, donde $\mu(t_j, \theta_i)$ es la función no lineal conocida; t_j es la variable independiente y θ_i el vector de parámetros que necesitan ser estimados. En un segundo escenario y a modo de comparación se asume que $Y_{ij} \sim Poisson(\mu(t_j, \theta_i))$.

La estructura de la media $\mu(t_j, \theta_i)$, la cual describe la relación entre el número acumulado de casos reportados y el tiempo, es el modelo de Richards [70], y puede ser expresado como sigue:

$$\mu(t_j, \theta_i) = \frac{\alpha_i}{[1 + k_i \cdot e^{-k_i \cdot \gamma_i (t_j - \eta_i)}]^{\frac{1}{k_i}}}, \quad i = 1, \dots, n \quad j = 1, \dots, J \quad (3.11)$$

Donde $\theta_i = (\alpha_i, k_i, \gamma_i, \eta_i)$ es un vector de parámetros de un área específica i a ser estimado. El parámetro α_i es el tamaño final de la epidemia, γ_i es la razón de crecimiento, η_i es el acmé de la epidemia y k_i es el exponente de desviación de la curva logística estándar.

La estimación de los parámetros de los modelos para cada área específica se realiza por los métodos expuestos en el capítulo 2, los cuales están implementados en el software R, utilizando las función **gnls** del paquete *nlme*.

3.3.2 Metodología propuesta: Modelo no lineal con efectos mixtos

El modelo no lineal con efectos mixtos correspondiente para el número acumulado de casos reportados Y_{ij} en el área de salud i en el tiempo t_j es

$$Y_{ij} = \frac{\alpha_i}{\left[1 + k_i \cdot e^{-k_i \cdot \gamma_i (t_j - \eta_i)}\right]^{\frac{1}{k_i}}} + \epsilon_{ij}, \quad (3.12)$$

El vector de parámetros para el área específica es modelado como:

$$\theta_i = X_i \theta + Z_i b_i \quad (3.13)$$

Donde θ es el vector de parámetros fijos, b_i es el vector de efectos aleatorios del área específica, X_i y Z_i son matrices de diseño conocido para los efectos fijos θ y los efectos aleatorios b_i , respectivamente:

$$X_i = Z_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \theta = \begin{pmatrix} \alpha \\ k \\ \gamma \\ \eta \end{pmatrix}, \quad b_i = \begin{pmatrix} b_{i1} \\ b_{i2} \\ b_{i3} \\ b_{i4} \end{pmatrix} \quad (3.14)$$

De la ecuación 3.13 sigue que el vector de parámetros de área específica puede ser expresada como

$$\begin{pmatrix} \alpha_i \\ k_i \\ \gamma_i \\ \eta_i \end{pmatrix} = \begin{pmatrix} \alpha + b_{i1} \\ k + b_{i2} \\ \gamma + b_{i3} \\ \eta + b_{i4} \end{pmatrix} \quad (3.15)$$

Se asume que los efectos aleatorios b_i están distribuidos normalmente como $b_i \sim N(0, \Psi)$ y el error dentro del grupo $\epsilon_{ij} \sim N(0, \sigma^2)$. Para representar la estructura de varianza-covarianza de los efectos aleatorios Ψ se utiliza una matriz general definida positiva.

Métodos de estimación de parámetros en un modelo no lineal con efectos mixtos

Diferentes métodos se han propuesto para estimar parámetros en un modelo no lineal con efectos mixtos. En este epígrafe se abordara un método basado en la función de verosimilitud

[6]. Primeramente, se considera un modelo no lineal con efectos mixtos de forma general:

$$y_{ij} = f(\phi_{ij}, v_{ij}) + \epsilon_{ij}, \quad i = 1, \dots, M \quad j = 1, \dots, n_i \quad (3.16)$$

donde M es el número de grupos o áreas en nuestro caso, n_i es el número de observaciones en el i -ésimo grupo, N el número total de observaciones, f es una función no lineal general diferenciable del vector de parámetros de un grupo específico ϕ_{ij} y de un vector de covariables v_{ij} , y ϵ_{ij} es el término error dentro del grupo, el cual está distribuido normalmente. El vector de parámetros de grupo específico ϕ_{ij} se modela como

$$\phi_{ij} = A_{ij}\theta + B_{ij}b_i, \quad b_i \sim N(0, \Psi) \quad (3.17)$$

donde θ es un vector p dimensional de efectos fijos y b_i es un vector q dimensional de efectos aleatorios asociado con el i -ésimo grupo (no varía con j) con matriz de varianza covarianza Ψ . Las matrices A_{ij} y B_{ij} son de dimensiones apropiadas y dependen del grupo y posiblemente de valores de algunas covariables en la observación j -ésima.

Como los efectos aleatorios son cantidades no observadas, la estimación de máxima verosimilitud en modelos con efectos mixtos está basada en la densidad marginal de las respuestas y , la cual para un modelo de un solo nivel de anidación, se calcula por

$$p(Y | \theta, \sigma^2, \Psi) = \int p(Y | b, \theta, \sigma^2) p(b | \Psi) db, \quad (3.18)$$

donde $p(Y | \theta, \sigma^2, \Psi)$ es la densidad marginal de Y , $p(Y | b, \theta, \sigma^2)$ es la densidad condicional de Y dado los efectos aleatorios b , y la distribución marginal de b es $p(b | \Psi)$. Para los modelos no lineales con efectos mixtos resulta conveniente expresar la matriz de varianza-covarianza de los efectos aleatorios en términos del factor de precisión Δ , donde $\Psi^{-1} = \sigma^{-2} \Delta^T \Delta$, ya que la densidad marginal de Y se puede expresar como

$$p(Y | \theta, \sigma^2, \Delta) = \frac{|\Delta|^M}{(2\pi\sigma^2)^{\frac{(N+Mq)}{2}}} \prod_{i=1}^M \int \exp \left\{ \frac{\|Y_i - f_i(\theta, b_i)\|^2 + \|\Delta b_i\|^2}{-2\sigma^2} \right\} db_i, \quad (3.19)$$

donde $f_i(\theta, b_i) = f_i[\phi_i(\theta, b_i), v_i]$

Como la función del modelo f es no lineal en los efectos aleatorios, la integral en (3.18) generalmente no puede ser determinada analíticamente, por lo que hay que recurrir a métodos aproximados de optimización numérica de la función de verosimilitud. Diferentes aproximaciones se han propuesto. Algunos de esos métodos consisten en tomar la expansión de Taylor de primer orden de la función f alrededor del valor esperado de los efectos aleatorios o alrededor de los modos condicionales (en Δ) de los efectos aleatorios.

En este epígrafe, se describe el método propuesto por Lindstrom y Bates [52]. Este método aproxima (3.18) por la verosimilitud del modelo lineal con efectos aleatorios y es llamado aproximación lineal de los efectos mixtos. Este es la base del algoritmo de estimación que está implementado en la función **nlme** del software R.

Algoritmo de Lindstrom y Bates Este algoritmo de estimación alterna entre dos pasos, un paso de mínimos cuadrados no lineales penalizado y un paso de efectos mixtos lineales (EML), que se describe a continuación:

En paso de mínimos cuadrados no lineales penalizado, las estimaciones actuales de Δ se mantienen fijas, y los modos condicionales de los efectos aleatorios b_i y las estimaciones condicionales de los efectos fijos θ se obtienen minimizando una función objetivo de mínimos cuadrados no lineales penalizadas

$$\sum_{i=1}^M \left[\|Y_i - f_i(\theta, b_i)\|^2 + \|\Delta b_i\|^2 \right] \quad (3.20)$$

El paso de efectos mixtos lineales actualiza la estimación de Δ basado en una expansión de Taylor de primer orden de la función del modelo f alrededor de las estimaciones actuales de θ y de los modos condicionales de los efectos aleatorios b_i , los cuales serán denotados por $\hat{\theta}^{(w)}$ y $\hat{b}_i^{(w)}$, respectivamente. Sean

$$\begin{aligned} \hat{X}_i^{(w)} &= \frac{\partial f_i}{\partial \theta^T} \Big|_{\hat{\theta}^{(w)}, \hat{b}_i^{(w)}}, & \hat{Z}_i^{(w)} &= \frac{\partial f_i}{\partial b_i^T} \Big|_{\hat{\theta}^{(w)}, \hat{b}_i^{(w)}} \\ \hat{w}_i^{(w)} &= y_i - f_i \left(\hat{\theta}^{(w)}, \hat{b}_i^{(w)} \right) + \hat{X}_i^{(w)} \hat{\theta}^{(w)} + \hat{Z}_i^{(w)} \hat{b}_i^{(w)}, \end{aligned} \quad (3.21)$$

La función de log-verosimilitud aproximada utilizada para estimar Δ es

$$l_{EML} = (\theta, \sigma^2, \Delta | y) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^M \left\{ \log |\Sigma_i(\Delta)| + \sigma^{-2} \left[\hat{w}_i^{(w)} - \hat{X}_i^{(w)} \theta \right]^T \Sigma_i^{-1}(\Delta) \left[\hat{w}_i^{(w)} - \hat{X}_i^{(w)} \theta \right] \right\}, \quad (3.22)$$

donde $\Sigma_i(\Delta) = I + \hat{Z}_i^{(w)} \Delta^{-1} \Delta \hat{Z}_i^{(w)T}$.

Un modelo lineal con efectos mixtos puede ser expresado

$$y_i = X_i \theta + Z_i b_i + \epsilon_i \quad i = 1, \dots, M \quad (3.23)$$

$$b_i \sim N(0, \Psi), \epsilon_i \sim N(0, \sigma^2 \mathbf{I}),$$

Una forma de obtener la función de verosimilitud se presenta a continuación. El modelo anterior (3.23) puede ser también expresado como

$$y_i = X_i \theta + Z_i b_i + \epsilon_i = X_i \theta + \epsilon_i^* \quad i = 1, \dots, M \quad (3.24)$$

donde $\epsilon_i^* = Z_i b_i + \epsilon_i$. Como los ϵ_i^* son la suma de dos vectores aleatorios independientes con distribución normal multivariada, ellos están distribuidos independientemente como vectores normales multivariados con media $\mathbf{0}$ y matriz de varianza-covarianza $\sigma^2 \Sigma_i$ donde $\Sigma_i = \mathbf{I} + \frac{Z_i \Psi Z_i^T}{\sigma^2}$. Entonces, de la ecuación (3.23) se observa que los y_i , son vectores aleatorios independientes con distribución normal multivariada con media $X_i \theta$ y matriz de varianza-covarianza $\sigma^2 \Sigma_i$. Es decir, la función de verosimilitud es

$$p(y_i | \theta, \Delta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n_i}{2}} \exp\left(\frac{(y_i - X_i \theta)^T \Sigma_i^{-1} (y_i - X_i \theta)}{-2\sigma^2}\right) |\Sigma_i|^{-\frac{1}{2}} \quad (3.25)$$

Para un valor dado de Δ , los valores de θ y σ^2 que maximizan la verosimilitud podrían escribirse como

$$\hat{\theta}(\Delta) = \left(\sum_{i=1}^M X_i^T \Sigma_i^{-1} X_i \right)^{-1} \sum_{i=1}^M X_i^T \Sigma_i^{-1} y_i$$

$$\hat{\sigma}^2(\Delta) = \frac{\sum_{i=1}^M \left(y_i - X_i \hat{\theta}(\Delta) \right)^T \Sigma_i^{-1} \left(y_i - X_i \hat{\theta}(\Delta) \right)}{N}, \quad (3.26)$$

Note, como la función de log-verosimilitud (3.22) es idéntica a la de un modelo lineal con efectos mixtos (3.25) en el cual el vector respuesta está dado por $\hat{w}^{(w)}$ y las matrices de diseños de los efectos fijos y aleatorios están dadas por $\hat{X}^{(w)}$ y $\hat{Z}^{(w)}$, respectivamente. Entonces,

usando los estimadores de (3.26), uno puede expresar los valores óptimos de θ y σ^2 como función de Δ y trabajar con la log-verosimilitud perfil de Δ , simplificando grandemente el problema de optimización.

Lindstron y Bates [52] también propusieron un método de estimación por máxima verosimilitud restringida para Δ , el cual consiste en reemplazar la log-verosimilitud en el paso de efectos lineales mixtos del algoritmo alternativo por la log-verosimilitud restringida

$$l_{EML}^R(\sigma^2, \Delta | y) = l_{EML}(\hat{\theta}(\Delta), \sigma^2, \Delta | y) - \frac{1}{2} \sum_{i=1}^M \log \left| \sigma^{-2} \hat{X}_i^{(w)T} \Sigma_i^{-1}(\Delta) \hat{X}_i^{(w)} \right|, \quad (3.27)$$

Note que, Como $\hat{X}_i^{(w)}$ depende de $\hat{\theta}^{(w)}$ y $\hat{b}_i^{(w)}$, cambios en el modelo de los efectos fijos o en el modelo de los efectos aleatorios implicaran cambios en el factor de penalidad para la log-verosimilitud restringida (3.27). Por lo tanto, las log-verosimilitudes restringidas de los modelos no lineales mixtos con diferentes modelos con efectos fijos o aleatorios no son comparables.

El algoritmo alterna entre los pasos de mínimos cuadrados no lineales penalizado y el paso de efectos mixtos lineales hasta que se alcance un criterio de convergencia. Tales algoritmos alternativos tienden a ser mas eficientes cuando las estimaciones de las componentes de la varianza-covarianza (Δ y σ^2) no están altamente correlacionadas con las estimaciones de los efectos fijos (θ).

Estrategia de construcción del modelo con efectos mixtos

En esta sección, se explica brevemente la estrategia que se debe seguir para obtener el modelo no lineal con efectos mixtos que mejor ayude a explicar la variabilidad de los parámetros epidemiológicos dentro de las áreas.

Un paso crucial en la construcción de modelos con efectos mixtos es decidir que coeficientes en el modelo necesitan efectos aleatorios para tener en cuenta la variación entre grupos y cuales deben ser tratados como efectos puramente fijos. Para ello, la estrategia de modelación es ajustar varios modelos teniendo en cuenta en cada modelo diferentes parámetros con efectos

aleatorios. Dentro de la estrategia, una alternativa es considerar en primer lugar un modelo sin efectos aleatorios. Luego, se deben utilizar todas las combinaciones de parámetros con efectos aleatorios. Aunque es importante mencionar, que este proceso de búsqueda se puede reducir si se utilizan los resultados del primer enfoque, tratado en la epígrafe anterior, ya que ese enfoque puede sugerir la estructura de los efectos aleatorios (parámetros con mayor variabilidad) y proporcionar valores iniciales de los parámetros. En esta alternativa se sugiere asumir que la matriz de varianza-covarianza tiene una estructura general definida positiva.

Otra estrategia alternativa de construcción del modelo es comenzar con un modelo con efectos aleatorios para todos los parámetros y entonces examinar los resultados del ajuste para decidir cuál de los efectos aleatorios debe ser eliminado del modelo. Un problema con este enfoque es que, cuando se asume una estructura general definida positiva para la matriz de varianza-covarianza de los efectos aleatorios, el número de parámetros a estimar crece con el cuadrado del número de efectos aleatorios. En casos donde el número de efectos aleatorios es relativamente grande al número de individuos (grupos, áreas), se recomienda generalmente utilizar inicialmente una estructura de matriz diagonal para la matriz de varianza-covarianza de los efectos aleatorios, para prevenir problemas de convergencia con un modelo sobre-parametrizado.

Durante este proceso, se sugiere observar en las salidas de los resultados, las estimaciones de las desviaciones estándar de los efectos aleatorios ya que desviaciones cerca de cero pueden sugerir que ese parámetro debe ser considerado como efecto fijo. Uno puede probar si cierto efecto aleatorio puede ser eliminado del modelo, realizando la prueba de razón de verosimilitud.

Otro aspecto importante a observar en los resultados, es la matriz de correlaciones estimadas de los efectos aleatorios porque puede sugerir la estructura de la matriz de varianza-covarianza. Por ejemplo, altas correlaciones entre dos efectos aleatorios comparados con otras correlaciones de otros efectos aleatorios sugiere que la matriz es no condicionada y que la estructura de los efectos aleatorios puede estar sobre-parametrizado, quizás sugiriendo una estructura diagonal en bloque. Un gráfico de diagnóstico útil para evaluar la sobre-parametrización, es el gráfico de dispersión de los efectos aleatorios estimados que aparecen en muchos software, en R se

puede utilizar la función **pairs** del paquete **nlme**. También se puede utilizar la prueba de razón de verosimilitud para probar si cierta estructura de matriz de varianza-covarianza debe ser considerada.

Finalmente, se debe chequear si los supuestos del modelo no lineal con efectos aleatorio parecen validos para el modelo ajustado a los datos. Para ellos el uso de métodos gráficos proporcionan las herramientas más útiles para evaluar esos supuestos. El gráfico de los errores estandarizados contra los valores ajustados, ayuda a evaluar el supuesto que los errores dentro del grupo son independientes e idénticamente distribuidos con distribución normal con media cero y varianza σ^2 . Si los residuos están distribuidos simétricamente alrededor de cero con una varianza aproximadamente constante, estos indicaría que no hay ninguna violación de este supuesto. Si hay un patrón diferente, no se cumpliría este supuesto y sería necesario, quizás, utilizar funciones de la varianza (implementadas en el paquete **nlme** de R) para modelar la heterocedasticidad. El gráfico de los errores estandarizados contra los cuantiles de la distribución normal ayuda a evaluar la violación del supuesto de normalidad para los errores dentro de los grupos.

La adecuación del modelo ajustado se debe visualizar mejor realizando un gráfico conjunto con los valores observados y ajustado, donde se observen las predicciones de la población (cuando los efectos aleatorios son ceros) y las predicciones entre grupos (usando los efectos aleatorios estimados). Un aspecto a tener en cuenta, es que el modelo final obtenido debe tener sentido práctico.

3.3.3 Aplicación de las metodologías

En esta sección se aplicarán ambos métodos a los datos. Primeramente, se discutirán los resultados obtenidos por ajustar modelos no lineales individuales para cada área por separado y luego se presentarán los resultados obtenidos utilizando el modelo con efectos mixtos. Los datos utilizados para esta aplicación pertenecen a todas las áreas de salud primaria del municipio Plaza, La Habana, donde se reportaron casos de dengue durante el brote del 2006 (Figura 5.6, Anexo 2).

Modelo para un área específica

El modelo no lineal de Richards en la ecuación 3.11 fue ajustado para cada área de salud por separado usando el procedimiento NLMIXED del SAS [76] y la función `nlsList` del paquete *nlsme* [64] del software R [66] (Ver códigos Anexo 7). Las tablas 3.6 y 3.7 muestran las estimaciones de los parámetros máximo verosímiles de cada área de salud específica (y los IC al 95%) cuando la respuesta se distribuye Poisson y Normal asintótica, respectivamente. Note que para cada una de las áreas específicas se obtuvo mejor bondad de ajuste cuando se asume que el número acumulado de casos reportados tiene distribución Normal asintótica. La figura 5.12 (Anexo 5) muestra los modelos ajustados individuales de las siete áreas primarias de salud cuando la variable respuesta se distribuye normalmente. Se observa una considerable variabilidad en las estimaciones de los parámetros.

Aunque los modelos para las áreas de salud específica ajustaron bien los datos, desde nuestro punto de vista, este abordaje tiene limitaciones. Utiliza 28 coeficientes para representar los perfiles individuales de los casos acumulados, indicados en la figura 5.12 (Anexo 5). Este abordaje es útil cuando se está interesado en modelar el comportamiento de un particular o conjunto fijo de áreas, pero no es adecuado cuando las áreas son consideradas como muestra de una (quizás hipotética) población y la inferencia debe centrarse en esta población.

Modelo no lineal con efectos mixtos

Como en este caso, el interés está en estimar el comportamiento medio de un área de salud en la población y la variabilidad entre y dentro de las áreas, se desarrolla un modelo con efectos mixtos. Un paso crucial en la construcción de un modelo con efectos mixtos es decidir cuales de los coeficientes en el modelo necesitan efectos aleatorios para tener en cuenta la variación entre las áreas y cuales puede ser tratado puramente como efectos fijos.

Siguiendo la estrategia de modelación, la tabla 3.8 muestra el criterio de información de Akaike (*AIC*) y el criterio Bayesiano de información (*BIC*) para los modelos no lineales mixtos para todas las combinaciones de los parámetros con efectos aleatorios asumiendo que la variable respuesta tiene distribución normal y Poisson respectivamente. En todos los casos

Tabla 3.6: Estimaciones de los parámetros del modelo de Richards para cada área de salud específica cuando la respuesta se distribuye Poisson.

Area	α (95%CI)	η (95%CI)	γ (95%CI)	k (95%CI)	R_0 (95%CI)	AIC
15 y 18	418.61 (397.45,439.76)	13.86 (13.42,14.30)	0.91 (0.45,1.37)	0.49 (0.12,0.86)	11.85 (0*,26.68)	146.5
19 de Abril	525.46 (502.92,547.99)	13.75 (13.39,14.11)	0.83 (0.54,1.11)	0.54 (0.24,0.82)	9.45 (2.20,16.69)	156.5
Corynthia	378.26 (357.96,398.55)	14.13 (13.54,14.72)	1.14 (0.11,2.18)	0.39 (0*,0.87)	22.27 (0*,85.01)	142.4
Moncada	430.16 (410.10,450.21)	14.84 (14.49,15.19)	0.75 (0.56,0.94)	0.68 (0.37,0.99)	7.63 (3.69,11.58)	159.9
Puentes Grandes	215.03 (201.97,228.09)	11.74 (10.83,12.66)	1.11 (0*,2.64)	0.38 (0*,1.08)	20.23 (0*,104.27)	148.3
Plaza	399.06 (382.94,415.18)	12.33 (11.98,12.68)	0.64 (0.51,0.77)	0.78 (0.48,1.08)	5.69 (3.72,7.66)	216.8
Rampa	454.47 (436.01,472.93)	14.40 (14.07,14.73)	0.63 (0.52,0.75)	0.94 (0.57,1.31)	5.59 (3.82,7.38)	165.7

Tabla 3.7: Estimaciones de los parámetros del modelo de Richards para cada área de salud específica cuando la respuesta tiene distribución Normal asintótica.

Area	α (95%CI)	η (95%CI)	γ (95%CI)	k (95%CI)	R_0 (95%CI)	AIC
15 y 18	420.05 (415.36,424.75)	13.76 (13.53,13.99)	1.07 (0.65,1.48)	0.40 (0.21,0.58)	18.22 (0*,38.94)	120.9
19 de Abril	524.14 (520.35,527.93)	13.87 (13.70,14.03)	0.73 (0.60,0.85)	0.64 (0.49,0.79)	7.17 (4.75,9.59)	126.5
Corynthia	381.09 (374.98,387.19)	14.79 (14.42,15.17)	0.59 (0.42,0.77)	0.92 (0.50,1.34)	5.00 (2.60,7.40)	129.2
Moncada	431.13 (425.10,437.16)	14.92 (14.63,15.22)	0.69 (0.49,0.89)	0.74 (0.44,1.04)	6.48 (3.02,9.94)	158.0
Puentes Grandes	215.32 (211.11,219.52)	12.58 (11.97,13.18)	0.51 (0.31,0.72)	1.08 (0.38,1.79)	4.03 (1.77,6.30)	135.4
Plaza	399.14 (387.35,410.94)	11.73 (11.39,12.07)	1.12 (0.56,1.68)	0.36 (0.14,0.57)	20.67 (0*,52.05)	202.1
Rampa	460.88 (456.82,464.93)	14.01 (13.81,14.21)	0.96 (0.68,1.25)	0.50 (0.31,0.68)	13.66 (3.15,24.17)	141.6

el parámetro α , el tamaño final de la epidemia, se consideró como efecto aleatorio por su gran variabilidad.

Tabla 3.8: El AIC y BIC para todas las combinaciones de parámetros con efectos aleatorios asumiendo que la variable respuesta tiene distribución Poisson y Normal.

Efectos aleatorios	AIC		BIC	
	Poisson	Normal	Poisson	Normal
-	3123.3	1730.0	3135.5	1745.3
α	1724.2	1434.7	1723.9	1453.0
α, η	1188.5	1135.6	1188.2	1160.0
α, γ	1395.8	1410.1	1395.5	1434.5
α, k	1441.6	1438.7	1441.3	1463.1
α, η, k	1191.4	1124.2	1191.1	1157.7
α, η, γ	1297.6	1127.8	1297.3	1161.3
α, k, γ	1355.5	1307.4	1355.1	1340.9
α, η, γ, k		1132.8		1177.9

Los menores valores de *AIC* y *BIC* fueron obtenidos para el modelo con los parámetros α , k y η como efectos aleatorios, es decir, considerando solo γ como parámetro fijo. Como el parámetro epidemiológico primario R_0 se obtiene de γ , entonces la estimación del parámetro R_0 , y por lo tanto la transmisión, sería la misma para cada área de salud si este modelo es considerado, es decir, $R_0 = 10.58(4.22, 16.94)$. Aunque el mejor ajuste se obtiene para este modelo, pudiera no tener sentido desde el punto de vista epidemiológico, en cuyo caso es mejor utilizar para inferencia el modelo cuando los parámetros con efectos aleatorios son α , η y γ , que tiene un *AIC* = 1127.8 y *BIC* = 1161.3 muy similar al del otro modelo. La tabla 5.6 (Anexo 5) muestra las estimaciones de los parámetros y la matriz de correlación de los efectos aleatorios para estos dos modelos obtenidas utilizando la función *nlme()* del paquete *nlme* [64] del software R [66].

La correlación estimada de 0.934 entre η y γ sugiere que la matriz de varianza-covarianza no está condicionada y que la estructura de los efectos aleatorios pudiera estar sobre-parametrizada. La matriz del gráfico de dispersión de los efectos aleatorios proporciona un diagnóstico útil

para evaluar los problemas de sobre-parametrización. La alineación casi perfecta entre los efectos aleatorios η y γ (Figura 5.13, Anexo 6) indica que el modelo está sobre-parametrizado. Las altas correlaciones entre los efectos aleatorios de η y γ y las bajas correlaciones entre estos efectos aleatorios y el efecto aleatorio de α sugieren que la matriz bloque-diagonal Ψ pudiera ser utilizada para representar la estructura de varianza-covarianza de los efectos aleatorios. Para probar si una matriz bloque-diagonal Ψ pudiera ser utilizada para representar la estructura de varianza-covarianza de los efectos aleatorios, se llevó a cabo una prueba de razón de verosimilitud (tabla 5.7, Anexo 5). El valor no significativo del *p-value* de la prueba de razón de verosimilitud y los valores más pequeños del *AIC* y *BIC* corroboran la estructura de varianza-covarianza de bloque-diagonal.

El gráfico de los errores estandarizados contra los valores ajustados correspondiente al modelo con α , η y γ como efectos aleatorios y con matriz de varianza-covarianza Ψ bloque-diagonal presentado en el panel de la izquierda de la figura 5.14 (Anexo 5), muestra que los residuos están distribuidos simétricamente alrededor del cero con una varianza aproximadamente constante. Esto no indica ninguna violación de los supuestos del modelo no lineal con efectos mixtos, excepto para algunas observaciones para el área 6 (Plaza). El gráfico de probabilidad normal de los errores estandarizados, mostrados en el panel de la derecha de la figura 5.14 no indica ninguna violación de los supuestos de normalidad para el error dentro los grupos.

Una evaluación final de la capacidad de este modelos está dado por el gráfico de las predicciones en la figura 3.1. Para comparación y para mostrar como los efectos individuales son tenidos en cuenta por el modelo no lineal con efectos mixtos, se presentan las predicciones de la población (correspondientes a los efectos aleatorios iguales a cero) y las predicciones dentro de los grupos (obtenidas utilizando los efectos aleatorios estimados). Note que las predicciones dentro de los grupos están en acuerdo con los casos acumulados observados, ilustrando que el modelo no lineal con efectos mixtos puede acomodar los efectos individuales.

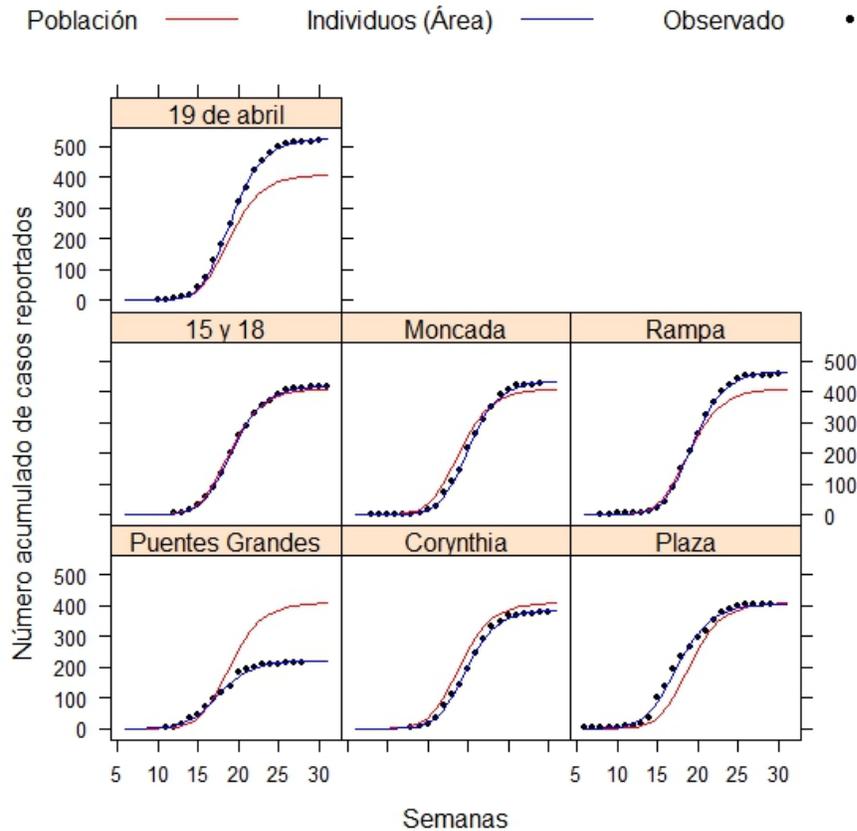


Figura 3.1: Predicciones de la población (población), predicciones dentro de los grupos (áreas), y número de casos acumulados observados (puntos) contra el tiempo en semanas, para el mejor modelo con efectos mixtos,

3.4 Conclusiones parciales

En este capítulo se proponen diferentes métodos estadísticos para la estimación de parámetros epidemiológicos en brotes de una onda pero cuando están coleccionados datos de casos acumulados para dos o más regiones.

En la situación que se quieren comparar dos brotes de dengue a través de las estimaciones de los parámetros epidemiológicos claves, se proponen diferentes alternativas desde el enfoque Frecuentista y Bayesiano basados en la extensión de un modelo no lineal (Modelo logístico 3P o modelo de Richards) que tenga en cuenta ambos brote y que permitan tener en cuenta la incertidumbre de modelos. En el enfoque Frecuentista se utiliza el promedio de modelos y en

el enfoque Bayesiano se utiliza el modelo Bayesiano jerárquico y un promedio Bayesiano de modelos aplicando selección Gibbs de variable.

La modelación de datos de brotes de dengue coleccionados para todas las áreas urbana de una particular región se lleva a cabo usualmente para cada área por separado. En este capítulo, se describen dos enfoques para estimar tres parámetros epidemiológicos claves: el acmé, el tamaño final de la epidemia y el número reproductivo básico. El primer enfoque consiste en ajustar un modelo no lineal individual para cada área por separado y en el segundo enfoque, la propuesta, se propone utilizar un modelo no lineal con efectos mixtos. Ambos enfoques fueron aplicados a los datos de siete áreas de salud del municipio Plaza, La Habana, durante el brote de dengue del 2006.

Para esta particular situación, es muy recomendado el segundo enfoque porque las áreas son consideradas como muestras de una población y no se ignora la variabilidad entre y dentro de las áreas. Sin embargo, el primer enfoque constituye una herramienta poderosa para la construcción del modelo del segundo enfoque porque las estimaciones individuales pueden sugerir el tipo de estructura de efectos aleatorios a utilizar y también proporcionar valores iniciales de los parámetros.

Capítulo 4

Modelación de brotes multiondas

4.1 Introducción

Hasta ahora se han discutido métodos estadísticos para la modelación de brotes de dengue de una onda. En este capítulo se abordan algunos métodos estadísticos para la modelación de brotes multiondas, es decir cuando hay más de una onda de infección.

Cuando hay más de una onda de infección, se propone una variación del modelo de Richards: El modelo de Richards multifases [35, 38–40], el cual hace distinción entre dos tipos de punto de retornos. Además del punto de inflexión de la curva inicial S que indica el primer punto de retorno o acmé de la epidemia donde termina el crecimiento inicial exponencial, existe un segundo punto de retorno en una curva epidémica multionda donde la razón de crecimiento del número de casos acumulados comienza a crecer nuevamente, significando el inicio de una nueva onda Hsieh [35, 38–40].

El procedimiento del modelo de Richards multi-etapas propuesto por Hsieh [35, 38–40] considera los segmentos sucesivos de casos acumulados en forma de S de forma separada, constituyendo esto desde nuestro punto de vista una desventaja. Una alternativa a esta limitación sería crear un modelo que tenga en cuenta todas las ondas epidémicas durante el brote conjuntamente y que permita la estimación de los parámetros de cada onda simultáneamente.

En este capítulo se propone una nueva metodología para la modelación de brotes multiondas, una variación del modelo de Richards multi-etapas propuesto por Hsieh [35, 38–40], el cual

tiene en cuenta todos los segmentos en forma de S simultáneamente para estimar el punto de retorno o acmé y el tamaño final de cada onda epidémica. Se presenta una propuesta desde el enfoque Frecuentista que hace uso de la verosimilitud perfil. La aplicación de la propuesta se muestra usando datos de la provincia La Habana del brote de dengue del 2001/2002 y además se muestra su uso para evaluar el impacto de las intervenciones ocurridas para el control del brote.

4.2 Modelación utilizando verosimilitud perfil

En esta sección se propone una nueva metodología para brotes multiondas desde el enfoque Frecuentista. Para una mejor comprensión, se presentará la metodología para brotes de dos ondas y luego se hará una generalización para brotes de tres ó más ondas.

4.2.1 Modelación de brotes de dos ondas

Sea Y_{tj} una variable aleatoria que describe el número acumulado de casos reportados en el tiempo t en la onda $j, j = (1, 2)$ y τ , el punto de retorno de segundo tipo ó el punto en el tiempo que indica el inicio de la segunda onda . Suponga que Y_{tj} tiene distribución normal asintótica con media $\mu(t, \theta)$ y varianza σ^2 , $Y_{tj} \sim N(\mu(t, \theta), \sigma^2)$ (a modo de comparación se puede también asumir que $Y_t \sim Poisson(\mu(t, \theta))$). Se supone además, que la segunda onda comienza una vez que haya terminado la primera onda. De acuerdo a este supuesto, la segunda onda tiene una asíntota inferior, la cuál coincide con la asíntota superior de la primera onda. Esta restricción se puede incluir en el modelo adicionando un nuevo parámetro α_0 a la estructura de media, que tiene en cuenta la presencia de una asíntota inferior para la segunda onda (Ver figura 4.1). Por lo que ahora la estructura de media se puede expresar como:

$$\mu(t, \theta) = \alpha_0 + \frac{\alpha - \alpha_0}{[1 + k \cdot e^{-\gamma k(t-\eta)}]^{\frac{1}{k}}} \quad (4.1)$$

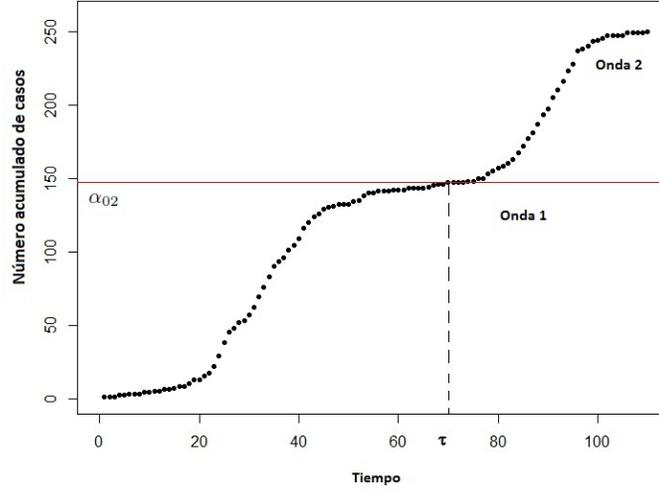


Figura 4.1: Número de casos acumulado para un brote de dos ondas, con una asíntota inferior para la segunda onda.

Donde el vector de parámetros $\theta = (\alpha_0, \alpha, \gamma, k, \eta)$ se define como: Sea

$$I_t = \begin{cases} 1 & t \leq \tau \\ 0 & t > \tau \end{cases} \quad (4.2)$$

Entonces

$$\begin{cases} \alpha_0 = \alpha_{02} \cdot (1 - I_t), \\ \alpha = \alpha_1 \cdot I_t + \alpha_2 \cdot (1 - I_t), \\ \gamma = \gamma_1 \cdot I_t + \gamma_2 \cdot (1 - I_t), \\ \eta = \eta_1 \cdot I_t + \eta_2 \cdot (1 - I_t), \\ k = k_1 \cdot I_t + k_2 \cdot (1 - I_t). \end{cases} \quad (4.3)$$

Donde α_{02} es la asíntota inferior de la segunda onda y los parámetros $\alpha_j, \gamma_j, \eta_j$ y k_j son los parámetros del modelo de Richards para cada una de las ondas $j = 1, 2$.

Otra manera de expresar la estructura de media es:

$$\mu(t, \theta) = \begin{cases} \mu(t, \theta_1) & t \leq \tau \\ \mu(t, \theta_2) & t > \tau \end{cases} \quad (4.4)$$

Donde las estructuras de media de cada onda sería:

$$\mu(t, \theta_j) = \alpha_{0j} + \frac{\alpha_j - \alpha_{0j}}{[1 + k_j \cdot e^{-\gamma_j \cdot k_j(t - \eta_j)}]^{\frac{1}{k_j}}} \quad (4.5)$$

Sabiendo que $\alpha_{01} = 0$ por ser la asíntota inferior de la primera onda.

Una vez definido el modelo, el principal interés está centrado en determinar el valor de τ . Para estimar el valor de τ , nos basaremos en la verosimilitud perfil. Para obtener la verosimilitud perfil para τ se fija $\tau = \tau_0$ y se maximiza el logaritmo de la verosimilitud de acuerdo a los otros parámetros ($\theta_1 = (\alpha_1, \gamma_1, \eta_1, k_1), \theta_2 = (\alpha_{02}, \alpha_2, \gamma_2, \eta_2, k_2)$). Se repite el proceso para diferentes valores de τ_0 en un intervalo dado. Los valores del logaritmo de la verosimilitud maximizada constituyen el logaritmo de la verosimilitud perfil, es decir

$$l_p(\tau_0, (\hat{\theta}_1, \hat{\theta}_2)) = \max_{(\theta_1, \theta_2)} l(\tau_0, (\theta_1, \theta_2)) \quad (4.6)$$

El estimador de τ , $\hat{\tau}$, es el valor de τ que maximiza la verosimilitud perfil. Con este valor $\hat{\tau}$ podemos estimar simultáneamente los parámetros θ_1 y θ_2 de cada una de las ondas epidémicas. Para construir un intervalo de confianza al $(1 - \alpha)100\%$ se utiliza un resultado de la distribución asintótica del estadístico de razón de verosimilitud:

$$D(\tau) = 2 \left\{ l(\hat{\tau}, (\hat{\theta}_1, \hat{\theta}_2)) - l_p(\tau, (\hat{\theta}_1, \hat{\theta}_2)) \right\} \sim \chi^2(1) \quad (4.7)$$

Por lo tanto, intervalo de confianza de τ al $(1 - \alpha)100\%$ se determina de la siguiente forma:

$$\begin{aligned} IC_{1-\alpha}(\tau) &= \left\{ \tau : D(\tau) \leq \chi_{1-\alpha}^2(1) \right\} \\ &= \left\{ \tau : 2 \left\{ l(\hat{\tau}, (\hat{\theta}_1, \hat{\theta}_2)) - l_p(\tau, (\hat{\theta}_1, \hat{\theta}_2)) \right\} \leq \chi_{1-\alpha}^2(1) \right\} \\ &= \left\{ \tau : -l_p(\tau, (\hat{\theta}_1, \hat{\theta}_2)) \leq \frac{\chi_{1-\alpha}^2(1)}{2} - l(\hat{\tau}, (\hat{\theta}_1, \hat{\theta}_2)) \right\} \\ &= \left\{ \tau : l_p(\tau, (\hat{\theta}_1, \hat{\theta}_2)) \geq l(\hat{\tau}, (\hat{\theta}_1, \hat{\theta}_2)) - \frac{\chi_{1-\alpha}^2(1)}{2} \right\} \end{aligned} \quad (4.8)$$

Es decir, los valores de τ para los cuales la verosimilitud perfil $l_p(\tau, (\hat{\theta}_1, \hat{\theta}_2)) \geq l(\hat{\tau}, (\hat{\theta}_1, \hat{\theta}_2)) - \frac{\chi_{1-\alpha}^2(1)}{2}$ tiene distribución normal asintótica con media $\mu(t, \theta)$ y varianza σ^2 , $Y_{tj} \sim N(\mu(t, \theta), \sigma^2)$ (a modo de comparación se puede también asumir que $Y_t \sim Poisson(\mu(t, \theta))$)

4.2.2 Modelación de brotes de tres ondas epidémicas ó más

En el caso de un brote epidémico de dengue con más de dos ondas epidémicas, se hace una generalización del modelo de dos ondas tratado previamente. Se supone igualmente que la variable Y_{tj} es una variable aleatoria que describe el número acumulado de casos reportados en el tiempo t en la onda $j, j = (1, 2, 3, \dots)$, la cual tiene distribución normal asintótica con media $\mu(t, \theta)$ y varianza σ^2 , $Y_{tj} \sim N(\mu(t, \theta), \sigma^2)$ (ó $Y_t \sim Poisson(\mu(t, \theta))$). Los puntos de retorno de segundo tipo τ_i serán, en cantidad, el número de ondas menos uno, por ejemplo, en brotes de dengue de tres ondas habrán dos puntos de retornos de segundo tipo, uno que indica el inicio de la segunda onda y otro que indica el inicio de la tercera onda epidémica. Además, se supone que la n -ésima onda comienza una vez que haya terminado la onda $n-1$. De acuerdo a este supuesto la asíntota superior de la onda $n-1$ coincide con la asíntota inferior de la n -ésima onda. Se ejemplifica como pudiera ser expresada la estructura de media para brotes de tres ondas. Para brotes de más de tres ondas sería de forma similar.

$$\mu(t, \theta) = \alpha_0 + \frac{\alpha - \alpha_0}{[1 + k \cdot e^{-\gamma k(t-\eta)}]^{\frac{1}{k}}} \quad (4.9)$$

Donde el vector de parámetros $\theta = (\alpha_0, \alpha, \gamma, k, \eta)$ se define como: Sea

$$I_{t1} = \begin{cases} 1 & t \leq \tau_1 \\ 0 & t > \tau_1 \end{cases} \quad I_{t2} = \begin{cases} 1 & \tau_1 < t \leq \tau_2 \\ 0 & t \leq \tau_1 \text{ or } t > \tau_2 \end{cases} \quad (4.10)$$

Entonces

$$\begin{cases} \alpha_0 = \alpha_{02} \cdot I_{t2} + \alpha_{03} \cdot (1 - I_{t1} - I_{t2}), \\ \alpha = \alpha_1 \cdot I_{t1} + \alpha_2 \cdot I_{t2} + \alpha_3 \cdot (1 - I_{t1} - I_{t2}), \\ \gamma = \gamma_1 \cdot I_{t1} + \gamma_2 \cdot I_{t2} + \gamma_3 \cdot (1 - I_{t1} - I_{t2}), \\ \eta = \eta_1 \cdot I_{t1} + \eta_2 \cdot I_{t2} + \eta_3 \cdot (1 - I_{t2} - I_{t3}), \\ k = k_1 \cdot I_{t1} + k_2 \cdot I_{t2} + k_3 \cdot (1 - I_{t1} - I_{t2}) \end{cases} \quad (4.11)$$

Donde α_{02} y α_{03} son las asíntotas inferiores de la segunda y tercera onda respectivamente y los parámetros $\alpha_j, \gamma_j, \eta_j$ y k_j son los parámetros del modelo de Richards para cada una de las ondas $j = 1, 2, 3$.

Una vez definido el modelo, el principal interés se centra en determinar los valores de τ_i (τ_1 y τ_2 , en el caso de tres ondas) y su selección está basada en los datos. Para ello se hace una

generalización del método de dos ondas. Un primer paso, sería obtener una estimación de τ_1 , $\hat{\tau}_1$, utilizando los datos de las dos primeras ondas. Por supuesto que quedaría la duda hasta que valor en el tiempo son los datos de la segunda onda porque esto es uno de los valores que se quiere estimar. Por eso se recomienda para estimar τ_1 utilizar los datos posteriores al acmé de la epidemia de la segunda onda y hacer este procedimiento para valores sucesivos en el tiempo mientras se observe convergencia en las estimaciones de los parámetros de cada una de las ondas. Por esta razón este procedimiento puede ser utilizado para hacer predicciones en tiempo real de los parámetros epidemiológicos.

Una vez obtenido el valor de $\hat{\tau}_1$ se necesita estimar el valor de τ_2 . Para estimar este valor también nos basaremos en la verosimilitud perfil condicionada dado que se conoce τ_1 y se utilizan todos los datos del brote de tres ondas (ó los datos hasta la tercera onda si tiene más ondas). Se construye la verosimilitud perfil para τ_2 fijando un valor para τ_2 y se maximiza el logaritmo de la verosimilitud, dado que τ_1 es conocido, de acuerdo a los parámetros θ_1 , θ_2 y θ_3 . Se repite el proceso para diferentes valores de τ_2 en un intervalo dado. Los valores del logaritmo de la verosimilitud maximizada constituyen el logaritmo de la verosimilitud perfil dado que τ_1 es conocido, es decir

$$l_p(\tau_2, (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) | \tau_1) = \max_{(\theta_1, \theta_2, \theta_3)} l(\tau_2, (\theta_1, \theta_2, \theta_3) | \tau_1) \quad (4.12)$$

El estimador de τ_2 , $\hat{\tau}_2$, es el valor de τ_2 que maximiza la verosimilitud perfil (4.12). Con este valor y el valor de $\hat{\tau}_1$ podemos estimar simultáneamente los parámetros de cada una de las tres ondas epidémicas. Los intervalos de confianza se pueden calcular de la misma forma que en el caso de dos ondas. Para un brote de dengue de más de tres ondas, se repite el mismo procedimiento.

4.3 Propuesta de procedimiento

En el epígrafe se propone un procedimiento alternativo para la estimación y predicción en tiempo real de los parámetros epidemiológicos: acmé, tamaño final de la epidemia y número reproductivo básico R_0 , para brotes de dengue multiondas. Este aspecto constituye también una novedad de la presente investigación. El procedimiento consiste en:

1. Ajustar el modelo de Richards al número de casos acumulados en semanas sucesivas asumiendo distribución Normal asintótica y de Poisson. Para brotes de una sola onda, la estimación de los parámetros convergerá en cuanto la trayectoria se aproxime a la capacidad de carga.
2. Si los parámetros estimados permanecen convergente hasta que no sean detectados más nuevos casos, el brote tiene solo una onda. Sin embargo, si las estimaciones comienzan a divergir desde un valor fijo, uno sabe que ha ocurrido el punto de retorno que denota el inicio de una segunda fase.
3. Estimar el punto de retorno que separa las dos fases en forma de S de una epidemia utilizando la verosimilitud perfil. Para ello es conveniente utilizar los datos posteriores al acmé de la epidemia de segunda onda.
4. Ajustar el modelo propuesto (extensión del modelo de Richards) a la curva de casos acumulados de las dos fases. Los parámetros estimados (de las dos ondas simultáneamente) convergerán nuevamente en cuanto la curva se aproxime a la capacidad de carga para la segunda fase.
5. Repetir los pasos 2-4 en los eventos que ocurran más fases, utilizando la verosimilitud perfil condicionada conocidos los puntos de retornos anteriores, hasta que termine el brote.
6. Hacer una validación de los escenarios considerados (análisis de estadísticos de bondad de ajuste, análisis de los residuos, etc.)

4.4 Ejemplos de aplicación

En esta sección se ilustra el uso de la alternativa presentada previamente para la modelación de brotes multiondas a los datos de los casos reportados para toda la provincia La Habana en el brote de dengue de 2001/2002 (Figura 5.7). Se considera solo el escenario donde se asume que el número de casos acumulados tiene distribución normal asintótica.

4.4.1 Utilizando verosimilitud perfil

Del gráfico de incidencia (Figura 5.7) se observa que el brote tiene tres ondas bastante bien definidas, por lo que se quiere estimar el número reproductivo básico, el acmé y el tamaño final de cada onda epidémica así como los puntos de retorno de segundo tipo que indican el inicio de la segunda y tercera onda epidémica. Como primer paso, se estima el valor de τ_1 utilizando la verosimilitud perfil. Para ello utilizamos los datos hasta la semana 31 del brote y se fija el valor de τ_1 para diferentes valores $\tau_1 = 21, 22, 23, 24$, y se calcula la log-verosimilitud perfil para cada uno de estos valores de τ_1 . En la tabla 4.1 se muestran estos resultados: De la tabla 4.1 se observa que la estimación puntual de τ_1 es $\hat{\tau}_1 = 23$.

Tabla 4.1: Verosimilitud perfil de τ_1 para diferentes valores.

τ_1	$l_p(\tau_1)$
21	-164.3156
22	-147.9143
23	-147.5707
24	-162.1596

Para determinar un intervalo de confianza al 95% se puede apoyar en la fórmula (4.8), es decir, $IC_{0.95}(\tau_1) = \left\{ \tau_1 : l_p(\tau_1, (\hat{\theta}_1, \hat{\theta}_2)) \geq l(\hat{\tau}_1, (\hat{\theta}_1, \hat{\theta}_2)) - \frac{\chi_{0.95}^2(1)}{2} \right\}$, pero como $l(\hat{\tau}_1, (\hat{\theta}_1, \hat{\theta}_2)) = -147.5707$ y $\chi_{0.95}^2(1) = 3.84$ entonces $IC_{0.95}(\tau_1) = \left\{ \tau_1 : l_p(\tau_1, (\hat{\theta}_1, \hat{\theta}_2)) \geq -147.5707 - \frac{3.84}{2} \right\} = \left\{ \tau_1 : l_p(\tau_1, (\hat{\theta}_1, \hat{\theta}_2)) \geq -149.4907 \right\}$. Para determinar el limite inferior y superior del intervalo de confianza se realizó un programa en R, del cual se obtuvo que la estimación de τ_1 es $\hat{\tau}_1 = 23$ $IC_{95\%}$ (21.9039, 23.1316). En el panel de la izquierda de la figura 4.2 se observa gráficamente este resultado.

Para estimar el otro punto de retorno τ_2 se utiliza la verosimilitud perfil condicionada a que $\tau_1 = 23$, para ello se utilizan todos los datos del brote y se fija el valor de τ_2 para diferentes valores $\tau_2 = 29, 30, 31, 32, 33$. Luego, se calcula la log-verosimilitud perfil dado que $\tau_1 = 23$ para cada uno de estos valores de τ_2 . En la tabla 4.2 se muestran estos resultados:

De la tabla 4.1 se observa que la estimación puntual de τ_2 es $\hat{\tau}_2 = 31$. Apoyados en la fórmula (4.8) se obtiene que $IC_{0.95}(\tau_2) = \left\{ \tau_2 : l_p(\tau_2, (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)) \geq -190.2491 \right\}$. De aquí se obtuvo

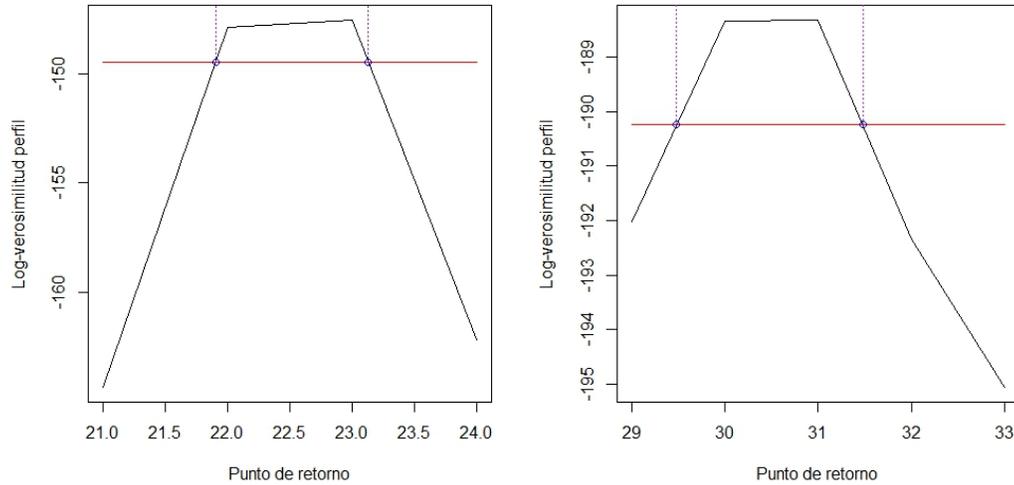


Figura 4.2: Log-verosimilitud perfil para los puntos de retornos. Panel de la izquierda (primer punto de retorno τ_1), panel de la derecha (segundo punto de retorno τ_2).

que la estimación de τ_2 es $\hat{\tau}_2 = 31$ $IC_{95\%}$ (29.4808, 31.4780). En el panel de la derecha de la figura 4.2 se observa gráficamente este resultado. Una vez estimado τ_1 y τ_2 se pueden estimar todos los parámetros de interés de cada una de las ondas epidémicas. En la tabla 4.3 se muestran las estimaciones de estos parámetros epidemiológicos claves.

Una evaluación final de la adecuación del modelo está dada en la gráfica de la figura 4.3, donde se visualizan los valores predichos acumulados y de incidencia obtenido por el modelo ajustado cuando $\tau_1 = 23$ y $\tau_2 = 31$. Este modelo será utilizado para estudiar el impacto de las medidas de control implementadas durante este brote.

4.4.2 Evaluación del impacto de una intervención

Para estudiar el impacto de las medidas de control estableceremos una correspondencia entre el modelo teórico de mejor ajuste utilizando verosimilitud perfil y las medidas de intervención y control implementadas.

Como parte de la prevención del dengue y durante el brote del 2001/2002 en La Habana, las acciones de vigilancia y de control fueron incrementándose en varias fases (Ver figura

Tabla 4.2: Verosimilitud perfil de τ_2 dado $\tau_1 = 23$ para diferentes valores.

τ_2	$l_p(\tau_2 \tau_1 = 23)$
29	-192.0156
30	-188.3417
31	-188.3291
32	-192.3456
33	-195.0486

Tabla 4.3: Estimaciones puntuales y error estándar de los parámetros epidemiológicos claves para cada una de las ondas epidémicas.

Parámetros	Estimaciones	Error estándar
α_1	5933.5	74.54
α_2	10812.9	67.23
α_3	12886.4	21.15
η_1	20.543	0.06328
η_2	28.201	0.3804
η_3	33.958	0.2553
R_{01}	2.9566	0.01531
R_{02}	1.1187	0.02398
R_{03}	2.9029	0.9130

4.4). A través de la vigilancia serológica a principio de junio se detecta el primer caso en la semana epidemiológica 22 del año 2001, primera semana de nuestro estudio. Tan pronto se detectó la transmisión de dengue, los médicos de la familia realizaron visitas regulares a las casas para buscar casos de dengue y comunicar mensajes sobre el control del dengue y del *Aedes aegypti*. Las acciones de control del vector aumentaron con fumigaciones en un radio de 100 m^2 alrededor de las casas de casos confirmados y comienzan a ser hospitalizados todos los casos de dengue confirmado. En la semana 19 se refuerzan las actividades de control del vector a nivel provincial (Ver figura 4.4). La frecuencia de los ciclos de inspección cambian de bimensual a mensual, se trabaja tanto en la eliminación de criaderos del vector como de las formas adultas del mosquito. A pesar de los cambios en la vigilancia, la intensificación de las

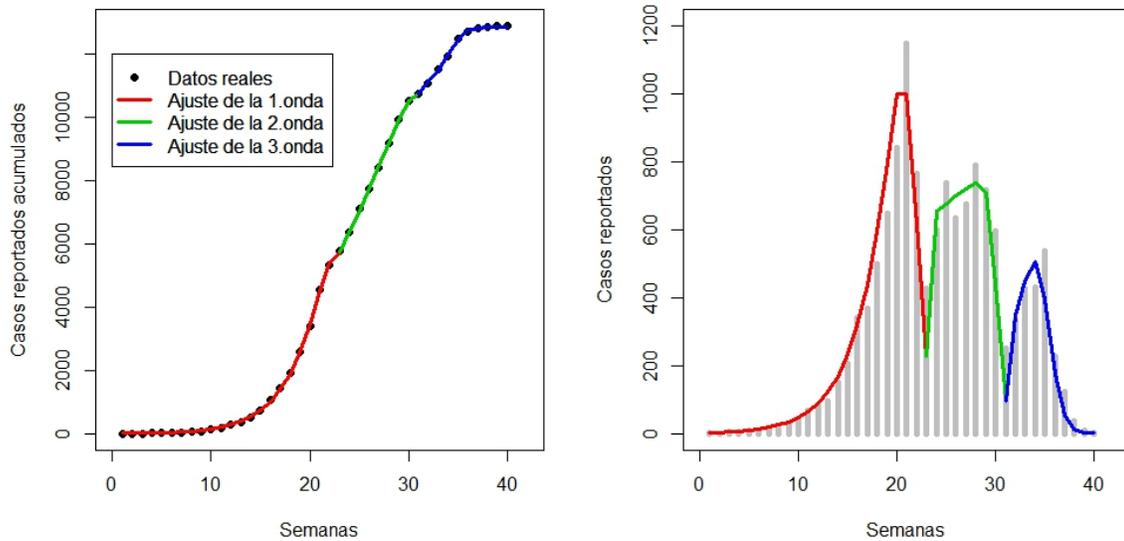


Figura 4.3: Modelo ajustado al número de casos acumulado (Panel de la izquierda) y a la incidencia (Panel de la derecha) para el brote de dengue multiondas de La Habana 2001/2002 utilizando verosimilitud perfil.

acciones de control del mosquito alrededor de los casos reportados y la política de aislamiento de los casos sospechosos de dengue en Hospitales, no es hasta la semana 21 (entre el 14 y 20 de octubre) que se obtiene el pico de la epidemia (descenso del reporte de casos), lo cual ocurre dos semanas después de ampliarse e intensificarse las acciones de control del vector en todos los municipios de la provincia. De acuerdo, al modelo, la estimación del acmé de la epidemia en esta primera onda fue $\hat{\eta}_1 = 20.54$, muy próximo a lo ocurrido.

Durante dos semanas el número de reportes de casos continuó decreciendo, sin embargo a partir de la semana 23 (entre el 4 y 10 de noviembre) comienza a aumentar el número de casos iniciándose una nueva onda epidémica, coincidiendo con la estimación del primer punto de retorno de segundo tipo $\hat{\tau}_1 = 23$ obtenida por el modelo. Este aumento en el número de casos reportados se debió en gran medida a que a partir de la semana 23 comenzó el sistema de vigilancia activo (Ver figura 4.4), iniciándose la búsqueda casa a casa de casos febriles para la detección de personas con sintomatología compatible con dengue a fin de mejorar la detección

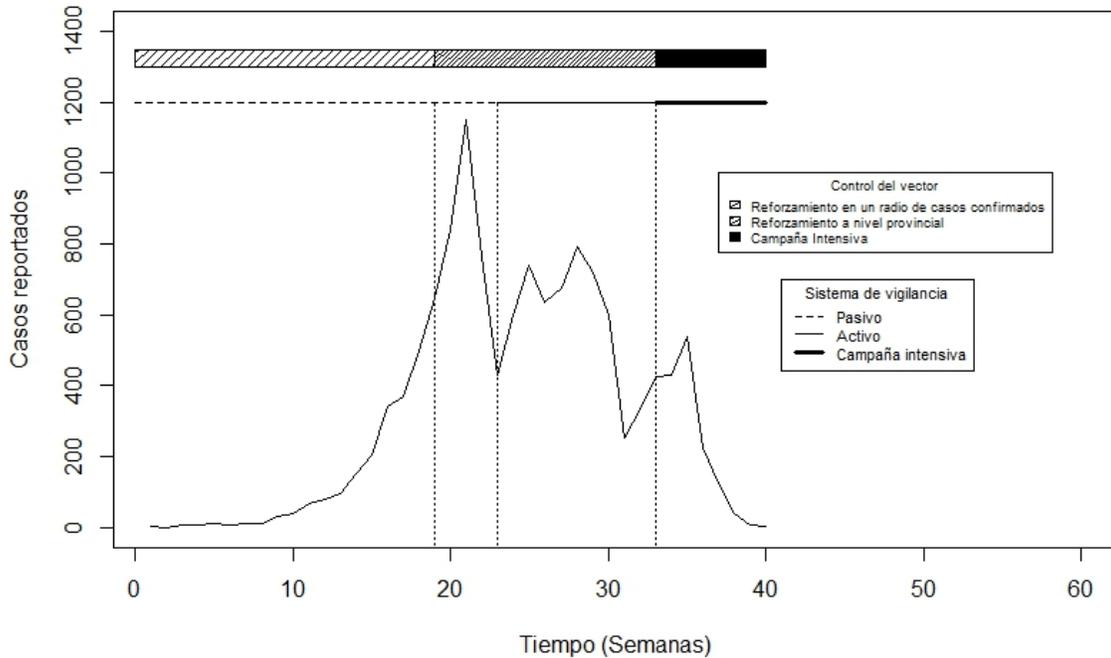


Figura 4.4: Sistema de vigilancia y control del vector durante el brote de La Habana 2001/2002.

y aislamiento de los casos en los hospitales especializados. Hseih y cols. [38], indican que este aumento se puede haber debido a que en la semana 24, Cuba fue azotada por el huracán Michelle.

Durante esta nueva onda infecciosa, el pico en el número de casos fue de 791 casos, que ocurrió en la semana 28 (entre el 2 y 8 de diciembre), estando en correspondencia con la estimación de $\hat{\eta}_2 = 28.201$ de nuestro modelo. A partir de esta semana el número de casos vuelve a descender como causa del sistema de vigilancia activo. Sin embargo, en la semana 31 comienza a crecer nuevamente el número de casos iniciándose una nueva onda epidémica, coincidiendo con la estimación del segundo punto de retorno $\hat{\tau}_2 = 31$ obtenida por el modelo. Este aumento a partir de la semana 31 se debió quizás al cambio en la política de hospitalización, los pacientes sospechosos de dengue fueron tratados en la casa y visitados diariamente por los médicos de familia. Este cambio en la política de aislamiento de los casos de dengue hace

que se incremente la posibilidad de contacto hombre enfermo-mosquito-hombre susceptible, aumentando posiblemente la transmisión y con esto el reporte de casos, aunque este nuevo aumento se puede haber debido al retraso del brote en algunos municipios principalmente en Habana Vieja [38].

Esta tercera onda epidémica tiene un pico en el número de casos en la semana 35, muy próxima a la estimación del acmé de la epidemia de esta onda obtenida por el modelo $\hat{\eta}_3 = 33.958$, dos semanas después de comenzada la campaña intensiva (Ver figura 4.4), dirigida a eliminar la transmisión con un enfoque multisectorial e interdisciplinario, con fuerte disciplina y control de las acciones y con la participación de todos los sectores políticos y sociales del país [62] [33]. Esta campaña intensiva revertió definitivamente el número de casos, permitiendo la eliminación del brote en marzo del 2002.

4.5 Conclusiones parciales

En este capítulo se propone una variante metodológica para la modelación de brotes multiondas con el objetivo de estimar los parámetros epidemiológicos que ayuden a evaluar las medidas de control implementadas durante un brote de dengue.

Desde nuestro punto de vista, el procedimiento del modelo de Richards multi-etapas propuesto por Hsieh [35, 38–40] presenta una desventaja ya que considera, para su estimación, los segmentos sucesivos en forma de S de forma separada. Por esa razón se propone un procedimiento alternativo que permite la estimación de los parámetros de cada onda simultáneamente. Esta propuesta se basa en la verosimilitud perfil.

En el capítulo se ilustra el uso de la alternativa para la modelación de brotes multiondas a los datos de los casos reportados para toda la provincia La Habana en el brote de dengue de 2001/2002 y se establece la correspondencia entre el modelo teórico de mejor ajuste utilizando la verosimilitud perfil y las medidas de intervención y control implementadas durante este brote.

En este enfoque de modelación se utilizó como estructura de la media el modelo de Richards.

En futuras investigaciones se extenderá esta metodología al uso de otros modelos no lineales para realizar inferencias multimodelos a través del promedio de modelos.

Capítulo 5

Desarrollo de Software: Paquete de R, *DengueRT*

Introducción

R [66] es una implementación flexible y poderosa de fuente libre del lenguaje S desarrollado por John Chambers y otros en los Laboratorios de la compañía Bells [20, 22] . A partir de 1995 comienza a ser distribuido gratuitamente bajo los términos de la licencia de GNU (Free Software Foundation) y desde entonces, el desarrollo de R ha sido un esfuerzo de colaboración internacional, con trabajo aportado por voluntarios. R puede ser extendido por medio de programas escritos por el usuario o mediante “bibliotecas” (*packages*) que pueden ser obtenidos vía Internet en forma gratuita [20, 25].

Uno de las fuerzas de R es la habilidad de compartir el software a través de los paquetes. Los paquetes dan a los usuarios una manera fiable, conveniente, y estandarizada de acceder a funciones de R, datos, y documentación [61]. Algunas de las principales razones para crear un paquete son i) los autores se verán obligados a pulir sus funciones, datos, códigos y, sobre todo, documentar todo el trabajo y dar ejemplos claros; ii) es el modo más elegante de compartir su trabajo, los posibles usuarios del paquete agradecerán el esfuerzo de mejora y documentación que requiere la creación de un paquete. El intercambio con los usuarios favorece la mejora del código; y iii) es la forma establecida para contribuir al crecimiento de R.

En este capítulo se presenta la implementación de los métodos, propuestos en el capítulo 2, en

un paquete de R, *DengueRT* [78]. Para ello, se hace una introducción a este paquete, donde se describen las principales funciones implementadas, los pasos para la creación del paquete con ayuda de **RStudio** [74] y el uso de la plataforma *R-forge* [22, 82]. La visualización de las funciones y gráficos del paquete se muestran con los datos del brote de dengue que ocurrió en dos áreas de salud del municipio Playa en La Habana, durante el brote de dengue del 2001/2002.

5.1 Introducción al paquete *DengueRT*

Una librería o paquete es una colección de objetos creados y organizados siguiendo un protocolo fijo que garantiza un soporte mínimo para el usuario así como la ausencia de errores (de sintaxis) en la programación. Los modelos de crecimientos no lineales y el método del promedio de modelos, discutidos en el capítulo 2, han sido implementados en el paquete de R, *DengueRT* [78].

El paquete *DengueRT* [78] está disponible en el *Comprehensive R Archive Network* (**CRAN**) en el sitio Web <https://cran.r-project.org/web/packages/DengueRT/>, desde mayo del 2016 (Anexo 8). Para trabajar apropiadamente el paquete *DengueRT* [78] requiere de los paquetes: *nlme* [64], *drc* [71] y *ggplot2* [22, 86].

5.1.1 Implementación de las funciones y datos

El primer paso para la elaboración de un paquete es la creación de todos los objetos de R: funciones, datos, etc. Lo primero que debe hacerse es programar las funciones y preparar los datos. El proceso de creación vigila que no hayan errores sintácticos pero no controla si hay errores lógicos. Para el paquete *DengueRT* [78] se crearon originalmente 16 funciones y se habilitaron dos bases de datos de brotes de dengue.

Las 16 funciones creadas fueron: *Richardsm*, *logistic3pm*, *SigmEmaxm*, *gompertz*, *weibullm*, *logistic5pm*, *weight*, *estimFSizeMA*, *estimTpointMA*, *CIFinalsize*, *CITurningpoint*, *allmodels*, *allmodelpredict*, *changetimeFSTP*, *summary* y *plot*. Sin embargo, algunas de estas funciones no son públicas para los usuarios.

Las primeras seis funciones: *Richardsm*, *logistic3pm*, *SigmEmaxm*, *gompertz*, *weibullm* y *logistic5pm* estiman los parámetros de interés, acmé y tamaño final de la epidemia, haciendo uso de la función *gnls* del paquete *nlme* [64], la cual ajusta modelos no lineales usando mínimos cuadrados generalizados. A pesar que estas seis funciones requieren para su funcionamiento de valores iniciales de los parámetros, con las funciones creadas se tuvieron en cuenta dos opciones. Una opción donde los usuarios pueden proporcionar los valores iniciales de los parámetros y otra, por defecto, de no utilizar valores iniciales para los parámetros ya que estos son proporcionados por funciones auto generadores de valores iniciales (*selfstart functions*) del paquete básico *stats* [66] (*SSlogis*, *SSgompertz* y *SSweibull*) y del paquete *drc* [71] (LL.4 y LL.5)

Las funciones *weight*, *estimFSizeMA*, *estimTpointMA*, *CIFinalsize* y *CIturningpoint* fueron programadas utilizando las fórmulas 2.32 - 2.36 del capítulo 2 para el cálculo de los pesos, de las estimaciones por promedio de modelos del tamaño final y acmé de la epidemia, así como los intervalos de confianza de estos parámetros respectivamente.

Todas estas funciones no están visibles en el paquete porque son ejecutadas por las funciones *allmodels* y *allmodelpredict* cuando son utilizadas para hacer un análisis retrospectivo o para hacer estimaciones en tiempo real. Las otras funciones públicas son las tres últimas *changetimeFSTP*, *summary* y *plot*. La tabla 5.1 presenta las principales funciones en el paquete y sus descripciones.

Las funciones *summary* y *plot* son funciones genéricas, lo que significa que ellas encierran un concepto genérico como un resumen y un gráfico respectivamente. Una función genérica realmente no hace ningún cálculo en sí. La implementación de una función genérica para una clase particular de objeto se llama método [87]. Estas funciones invocan métodos particulares de la clase S3 *dengue*. Las clases S3 fueron creadas por John Chambers, uno de los creadores originales del idioma de S. Esta clase se introdujo en versión 3 de S, de ahí su nombre. En particular, cada una de las clases S3 tiene una lista y un nombre. Los elementos de la lista son las variables del miembro de la clase. En el paquete creado se creó la clase S3 denominada *dengue*, una lista con las salidas de los objetos *allmodels*, *allmodelpredict* y

Tabla 5.1: Las principales funciones del paquete *DengueRT*.

Función	Descripción
<code>allmodels()</code>	Ofrece la estimación de parámetros para cada modelo incorporado y para el promedio de modelos del tamaño final y acmé del brote epidémico.
<code>allmodelpredict()</code>	Ofrece predicciones en tiempo real de estos parámetros para cada modelo incorporado y para el promedio de modelos.
<code>changetimeFSTP()</code>	Ofrece los cambios en el tiempo de las estimaciones del tamaño final y el acmé de la epidemia para cada modelo incorporado y para el promedio de modelos.
<code>summary()</code>	Función genérica que produce resultados resúmenes de las funciones: <code>allmodels()</code> , <code>allmodelpredict()</code> y <code>changetimeFSTP()</code> .
<code>plot()</code>	Función genérica que produce gráfico para la visualización de los resultados de las funciones: <code>allmodels()</code> , <code>allmodelpredict()</code> y <code>changetimeFSTP()</code>

changetimeFSTP. Es decir, estas dos funciones visualizan los resultados y produce resúmenes de los objetos *allmodels*, *allmodelpredict* y *changetimeFSTP* de la clase S3 *dengue*. Una descripción completa de este paquete se puede encontrar en su manual y ficheros ayuda (Anexo 9).

Las bases de datos *dengueoutbreak1* y *dengueoutbreak2* se encuentran disponibles en el paquete. Ellas contienen el número de casos reportados por semanas de los brotes de dengue de una onda del 2001-2002 en las áreas de salud “26 de julio” y “J.R. Ramirez” del municipio Playa de la provincia La Habana, Cuba. Estas bases de datos fueron proporcionadas por el Instituto de Medicina Tropical “Pedro Kouri” (IPK). Estos conjuntos de datos fueron introducidos en el paquete con la instrucción `save()`. Los nombres de los objetos se ponen como primer argumento. En el argumento “file” se debe especificar la carpeta donde dejaremos los objetos. Por ejemplo, para añadir el data.frame “dengueoutbreak1”, se grabará como “Rda” con la instrucción, `save(dengueoutbreak1, file=“C:/.../DengueRT/data/dengueoutbreak1.Rda”)`.

5.1.2 Creación del paquete *DengueRT*

El proceso para crear paquetes de R se describe con todo detalle en los manuales [46, 67] del R Core Team y que se pueden ver en el **CRAN**. Aunque se puede crear un paquete

únicamente desde la consola de R y la consola del sistema, es preferible utilizar un programa como **RStudio** [74] que facilitará algunas tareas.

R fue diseñado en un entorno *Unix* que incluye un conjunto de herramientas como compiladores, utilidades de programación y rutinas de consola. Windows carece de estos componentes, de forma que se deben descargar e instalar algunos programas de software libre equivalentes a los de *Unix*. Las componentes necesarias serían 1) un conjunto mínimo de utilidades tipo *Unix*, las llamadas *Rtools* de Brian Ripley y que ahora mantiene Duncan Murdoch; 2) Algunos compiladores *GNU* reunidos en MinGW-w64 (necesarios únicamente si su paquete contiene código C, Fortran o C++); 3) El compilador de la ayuda HTML de Microsoft y 4) Una versión del procesador de textos *TeX*. La obtención de las dos primeras componentes se hace directamente del **CRAN**, seleccionando el ejecutable **RtoolsXX.exe** más actual.

Una vez creadas todas las funciones y datos del paquete, el próximo paso es la creación de la estructura de las carpetas y archivos. Para crear un nuevo paquete en **RStudio** se utiliza la instrucción *Create Project* desde el menú *Project* o desde la barra de herramientas. Se debe seleccionar el *Type: Package*, dar un nombre al paquete que será también el nombre del proyecto y el de la carpeta y decirle donde crearlos (carpeta y proyecto). También hay que añadir el archivo fuente con las funciones y datos que queremos incorporar al paquete. Así, cuando pulsemos el botón *Create Project*, se creará toda la estructura de archivos necesaria dentro de la carpeta especificada.

El contenido de la carpeta creada es el siguiente:

1. En su nivel superior se encuentran los archivos de definición y de información como:
 - **Read-and-delete-me**: Indicaciones para la creación del paquete.
 - **DESCRIPTION**: Describe las características principales del paquete.
 - **NAMESPACE**: Carga los paquetes necesarios y exporta las funciones.
2. Algunas carpetas como:
 - **R**: funciones del paquete.

- **man** documentación o ayuda.
- **src** código en un lenguaje de bajo nivel (opcional).
- **data** bases de datos (opcional)

El próximo paso es el de editar los archivos de configuración y documentación del paquete. El primero y fundamental es el archivo **DESCRIPTION**. Para ello debemos rellenar diferentes campos, entre los que se encuentran: *Package* (nombre), *Type* (package), *Title* (título, una sola línea), *Description* (descripción del paquete, un párrafo corto), *Author* (autor, pueden ser varios), *Maintainer* (la persona que mantiene el paquete con su correo electrónico), *Version* (debe relacionar la versión actual), **Date** (opcional, debe relacionar la fecha de la versión actual), *License* (una cadena tal como “GPL-3” refiriéndose al fichero del subdirectorio `share/licenses` del paquete de R), *LazyData* (Si se tienen objetos de datos y se pone True, esto eliminará la necesidad de usar *data()* antes de usar el conjunto de datos), *Depends* (lista de paquetes que se requieren para que corra el paquete, puede incluir detalles de la versión requerida) y otros campos opcionales, para detalles ver el manual [67]. El archivo **DESCRIPTION**, del paquete creado, se muestra en el Anexo 8.

Otro archivo importante es el **NAMESPACE**. Sirve para cargar los paquetes o librerías que necesita nuestro paquete y para indicar qué funciones haremos públicas (con ayuda). Dentro de las entradas más importantes en este archivo están: *export* (se ponen las funciones que queremos hacer visibles), *import* (lista de los paquetes requeridos para que corra el paquete), *importFrom* (se utiliza para importar solo algunas funciones de paquetes) y *S3method* (se utiliza si se utilizan funciones genéricas indicando el método S3 con la función y la clase). El archivo **NAMESPACE** del paquete *DengueRT* [78], se muestra también en el Anexo 8 .

El siguiente paso, la documentación de las funciones, es uno de los más engorrosos, pero a la vez uno de los más importantes ya que si nuestra documentación es confusa el paquete no se podrá utilizar correctamente. La subcarpeta **man** contiene los archivos de documentación (Rd) de cada uno de los objetos utilizables del paquete. Para cada función de la carpeta **R**, debemos crear un archivo de documentación (con la extensión Rd) en la carpeta **man**. R

utiliza un lenguaje especial para crear estos archivos de documentación “Rd”. La instalación de R se encargará de trasladar este lenguaje genérico Rd a archivos HTML, de texto o LaTeX, según se necesite. Este lenguaje de documentación tiene una semejanza evidente con el LaTeX.

Con la utilización del **RStudio** [74], ya tenemos una plantilla “Rd” para cada función de nuestro paquete. Ahora debemos editar y rellenar la información requerida. El lenguaje “Rd” consiste en una serie de instrucciones con sus correspondientes argumentos. Cada instrucción es de la forma: `\commando{arg}`. El archivo “Rd” para una función contiene los puntos obligatorios `\name{}` (nombre del fichero ayuda), `\alias{}` (tópico que usualmente incluye el nombre, se puede utilizar el mismo documento de ayuda para varios objetos), `\title{}` (corta descripción del tópico), `\description{}` (algunas líneas describiendo el tópico), `\usage{}` (la sintaxis exacta de la función, mostrando sus argumentos) en la cabecera y `\keyword{}` (palabras claves, elegir de una lista en el directorio `/doc` de R) en el pie, además de los puntos opcionales como `\arguments{}` (descripción de cada argumento), `\value{}` (descripción de los valores de salida), `\details{}` (descripción detallada y precisa de lo que hace la función), `\references{}` (referencias a la literatura), `\seealso{}` (indicadores a los objetos de R relacionados), `\examples{}` (ejemplos de cómo usar la función, deberían correr correctamente y no tomar demasiado tiempo) en el cuerpo del archivo.

Además de los puntos obligatorios, se recomienda añadir las secciones `\arguments{}`, `\value{}`, `\author{}` y `\examples{}`. En todo caso la recomendación es empezar con un mínimo de puntos y añadir los que deseemos después de depurar de errores el archivo. Los mensajes de error de R para los archivos de documentación pueden ser muy crípticos. Por ello es conveniente tener mucho cuidado en los detalles de su redacción y se recomienda crear estos archivos de forma incremental y trabajar con una única función cada vez. Los ejemplos deben ser ejecutables. R comprobará durante la instalación que los ejemplos funcionan. Las páginas de ayuda para conjuntos de datos siguen una estructura similar y contienen los items `\name{}`, `\alias{}`, `\doctype{}`, `\title{}`, `\description{}`, `\usage{}`, `\format{}` (descripción del objeto, lista o marco de datos), `\source{}` (detalles de la fuente original), `\examples{}` y `\keyword{}`, donde `\doctype{}` viene definido como `data` y `\keyword{}` como `datasets`. Las

ayudas de cada una de las funciones y de los conjuntos de datos del paquete se encuentran en el Anexo 9.

Una vez creada la documentación de las funciones y datos el próximo paso es chequear el paquete, es decir, verificar que no haya errores de sintaxis o no se generen advertencias (*warnings*). **RStudio** [74] dispone de una pestaña específica para chequear el paquete. Si pulsamos el icono **Check** se chequea el paquete y veremos los errores o advertencias. En **RStudio** [74] también es posible chequear cada archivo Rd de forma individual ya que dispone de un botón para ver el resultado en HTML. Corregidos los errores y todas las advertencias que podamos, el siguiente paso es construir el paquete que lo podemos hacer con el icono **Build & Reload** que compilará el paquete generando un archivo con extensión .tar.gz (nombredelpaquete.tar.gz) listo para su instalación. La instalación del paquete con el archivo .tar no tendrá ninguna dificultad en linux. Sin embargo para instalar el paquete en otro ordenador Windows será mejor convertirlo al formato .zip. Con el icono **More** se despliega un menú que permite crear el paquete en la versión .zip en Windows con la opción **Build Binary Package**.

5.1.3 Uso de la plataforma *R-forge*

Una herramienta muy útil para la creación de paquetes de R es el **R-forge** [22, 82]. **R-forge** [82] proporciona un conjunto de herramientas para el manejo de códigos fuentes y varias características basadas en web. Este tiene como objetivo proporcionar una plataforma para el desarrollo colaborativo de paquetes de R, software relacionados con R u otros proyectos. Los desarrolladores de software usualmente utilizan herramientas como Subversión (SVN) o Sistema de Versiones Concurrentes (CVS) para manejar los códigos base eficientemente.

Para la creación de paquetes es conveniente utilizar **R-forge** [22, 82] ya que el código del paquete queda accesible remotamente, permitiendo la colaboración en su desarrollo de distintos ordenadores, por distintas personas incluso desde ubicaciones distintas; además el código del paquete queda gestionado bajo Subversión. Versiones, conflictos, ediciones simultáneas, actualizaciones, etc. Se gestionan automáticamente. **R-forge** [82] también comprueba diariamente el paquete en tres plataformas distintas y bajo dos versiones diferentes de R (producción

y desarrollo). Otras ventajas de esta herramienta es que aunque el paquete no esté alojado en los servidores del **CRAN**, queda disponible para su instalación por parte de cualquier usuario de R, descargándolo del enlace correspondiente. Además, con esta herramienta se facilita el proceso de subir el paquete al **CRAN** con solo completar un pequeño formulario.

Para registrarse como nuevo usuario de *R-forge*, se debe pulsar el enlace *New Account* en el parte superior derecha del navegador de Windows en el sitio <http://R-Forge.R-project.org>. Se debe completar el formulario (Hay algunas descripciones y sugerencias para cada campo en este sitio) y se debe pulsar después *Submit*. Se recibirá un correo electrónico con un URL a la dirección electrónica especificada. Después de pulsar este enlace su cuenta se verifica y se puede registrar en el sitio Web. La pestaña *My Page* es la página más importante de en *R-forge*. Ahí se puede configurar su cuenta, ver los miembros del proyecto y ver los aspectos que le han sido asignados.

En *R-forge* todo está organizado a través de proyectos. En estos proyectos es posible organizar uno o más paquetes y otros materiales relacionados con R. Una vez registrado en este sitio se puede comenzar su propio proyecto. Para registrar un nuevo proyecto, se debe ir al sitio Web de *R-forge*, registrarse e ir a la sección *My Page*. Pulse en el enlace *Register Project* en el menú de la parte superior de esa página. Se debe llenar el formulario y someter su proyecto. Después que el proyecto sea aprobado por los administradores de *R-forge* será notificado a través del correo electrónico y será capaz de comenzar con su proyecto en *R-forge*.

R-forge utiliza Subversión (SVN, <http://subversion.tigris.org>) para la gestión de códigos fuentes. Se necesita de un cliente SVN (por ejemplo *Tortoise SVN* en ordenadores Windows, <http://tortoisesvn.tigris.org>) para aprovechar completamente el repositorio SVN. Por razones de seguridad se utiliza una cubierta segura (*secure shell, SSH*) abriendo un túnel para las cuentas de los desarrolladores, lo que significa que todo tráfico en la red está encriptado. Esto significa que se tiene que configurar la máquina correctamente: 1) Descargar e instalar el último cliente *Tortoise SVN* desde <http://tortoisesvn.net/downloads>; 2) Es suficiente utilizar la autenticación de claves para escribir al repositorio del proyecto (si se decide

hacer esto , ir al paso 5), pero por conveniencia es probablemente mejor generar una llave SSH y subir la llave pública a **R-forge** y esto se hace: i) Descargando y ejecutando *putty-gen.exe* en <http://www.chiark.greenend.org.uk/~sgtatham/putty/>. ii) Se deja marcado *ssh-2rsa* y se debe pulsar **generate**, iii) Se guarda la llave privada utilizando el botón correspondiente, iv) Se marca el texto en el campo de texto **Public key for pasting into OpenSSH authorizedkeys files**, v) Se copia y se pega en la configuración de la información de su cuenta **shell**: Ir a <http://R-forge.R-project.org> y registrarse, ir a **My Page** y luego pulsar en **Account maintenance**, en parte final de esta página pulse en **edit key** en la ventana **Shell Account Information** (debe ser administrador o miembro del proyecto para hacer esto. 3) Ahora tiene que esperar hasta la próxima hora complete, las llaves son activadas una vez solo en una hora; 4) Ahora se necesita de un agente de autenticación como el *pageant.exe*. Cargue su llave privada con *pageant.exe* (clic derecho en el icono de bandeja del *pageant* y luego **add key**); 5) Finalmente chequee el repositorio utilizando el URL dado en el sitio Web del proyecto (pestaña **SCM**) dentro la cuenta del desarrollador con **Tortoise SVN** (si este último paso falla por alguna razón, espere una hora e inténtelo nuevamente).

Para construir un paquete, simplemente coloque su paquete en el directorio **pkg/** en su repositorio **SVN** y lo chequea. Usualmente este directorio contiene el paquete de R con el fichero usual **DESCRIPTION** y los directorios **R/**, **man/**, **data/**. Los paquetes son chequeados automáticamente cada noche (Horario de Europa central), contruidos y chequeado en varias plataformas. Todos los paquetes chequeado con éxito aparecen en la página del proyecto en la pestaña **R packages**. Los miembros del proyecto también tienen acceso a los registro de chequeo y construcción en todas las plataformas. Los paquetes pueden ser sometidos o enviados al **CRAN** a través de esta página Web de su proyecto, solo que debe ser administrador. El paquete **DengueRT**, antes de ser publicado en el **CRAN**, se creó como el proyecto de **R-forge: DengueParmEst**.

5.2 Visualización de las funciones del paquete *DengueRT*

En esta sección ilustramos el uso del paquete *DengueRT* [78] en el conjunto de datos relacionados con el brote de dengue de una onda ocurrido en las áreas de salud “26 de julio” y “J.R. Ramirez” del municipio Playa de la provincia La Habana, Cuba. Se muestran dos ejemplos de aplicación, el primero para estimar el acmé y el tamaño final de la epidemia permitiendo estudiar retrospectivamente el significado de varios eventos que ocurrieron en diferentes momentos en el área de salud y evaluar el impacto de las medidas de intervención y control implementadas durante el brote de dengue y el segundo para hacer predicciones en tiempo real de estos parámetros de interés.

5.2.1 Estimación de parámetros. Ejemplo de aplicación en dengue

Usualmente el primer paso para el uso del paquete de R, *DengueRT* [78], es definir el vector incidencia y el vector tiempo. Estos son vectores de igual longitud que especifican la incidencia (número de casos reportados por unidad de tiempo) y el intervalo de tiempo (desde el inicio del brote). Las bases de datos `dengueoutbreak1` y `dengueoutbreak2` contienen los datos de las dos áreas de salud, respectivamente.

```
> library(DengueRT)
# Datos del área de salud 1
> data(dengueoutbreak1)
# Datos del área de salud 2
> data(dengueoutbreak2)
> head(dengueoutbreak1)
  Incidence Time
1          1    1
2          0    2
3          2    3
4          5    4
5          2    5
6          2    6
> head(dengueoutbreak2)
  Incidence Time
```

```

1      5      1
2      4      2
3      4      3
4      3      4
5      7      5
6      1      6

```

En un segundo paso, se debe decidir cuál modelo no lineal utilizar para estimar los parámetros, si será un modelo individual de los modelos incorporados al paquete o si serán utilizados todos los modelos no lineales. La función `allmodels()` proporciona la estimación de los parámetros tamaño final y el acmé de la epidemia para las dos opciones y, en el caso que sean utilizado todos los modelos, ofrece el AIC de cada modelo, los pesos de cada modelo, la predicción de la incidencia y de los casos acumulados. Para producir un resumen de estos resultados se utiliza la función `summary()`. Se mostrará su uso, primero con un modelo individual y luego con todos los modelos no lineales incorporados al paquete.

```

# Estimación de los parámetros utilizando el modelo de Richards
# (Área de salud 1)
> h1 <-allmodels(dengueoutbreak1$Incidence,dengueoutbreak1$Time, model="Richards")
> summary(h1)
# Salida 1
Richards model
AIC
214.0607
Parameter estimate
      Value Std.Error  t-value  p-value
alpha 375.0832264 1.14702012 327.006667 5.761598e-58
k      2.2914315 0.23076559  9.929693 2.693529e-11
gamma  0.3284212 0.01383574 23.737170 6.951558e-22
eta    15.8038168 0.12280456 128.690793 5.126656e-45

Final size estimate:
      lower  est.  upper
372.7468 375.0832 377.4196

```

Turning point estimate:

```
  lower    est.    upper
15.55367 15.80382 16.05396
```

En el caso que sea de interés las estimaciones del promedio de modelo se utiliza la función `allmodels()` de la siguiente forma:

```
# Estimación de los parámetros utilizando todos los modelos no lineales
#(Área de salud 2)
> h2 <-allmodels(dengueoutbreak2$Incidence,dengueoutbreak2$Time,model="all")
> summary(h2)
```

La estimación de los parámetros se muestra en el siguiente panel.

```
# Salida 2
AIC
      Richards      3P Logistic Sigmoidal Emax      Gompertz      Weibull
      234.7269      240.6719      222.4572      238.4828      250.1002
5P Logistic
      220.7851
Model weights
      Richards      3P Logistic Sigmoidal Emax      Gompertz      Weibull
      6.54415e-04      3.34912e-05      3.02130e-01      1.00063e-04      3.00323e-07
5P Logistic
      6.97081e-01
Model specific and model average estimate of the final size
      lower    est.    upper
Richards      339.8966 343.8305 347.7645
3P logistic      341.7937 346.0720 350.3503
SigmEmax      344.4335 348.0050 351.5764
Gompertz      343.5955 347.9745 352.3536
Weibull      347.2308 353.6570 360.0832
5P logistic      342.6305 346.2744 349.9184
Model averaged 341.8111 346.7956 351.7801
Model specific and model average estimate of the turning point
```

	lower	est.	upper
Richards	13.29229	13.74728	14.20227
3P logistic	12.98689	13.16641	13.34593
SigmEmax	13.20638	13.35525	13.50413
Gompertz	12.14338	12.33392	12.52446
Weibull	12.15957	12.36931	12.57904
5P logistic	13.22684	13.37713	13.52742
Model averaged	13.17956	13.37064	13.56172

Estos resultados pueden ser visualizados utilizando la función `plot()`. Cuando el argumento de esta función es un objeto `allmodels`, utilizando la opción gráfica, `which`, la función traza la curva epidémica acumulada (`which=1`) o la curva epidémica de incidencia (`which=2`) con los modelos ajustados, las estimaciones puntuales y por intervalo de confianza al 95% del tamaño final de la epidemia (`which=3`) o del acmé de la epidemia (`which=4`). Se muestra su uso debajo.

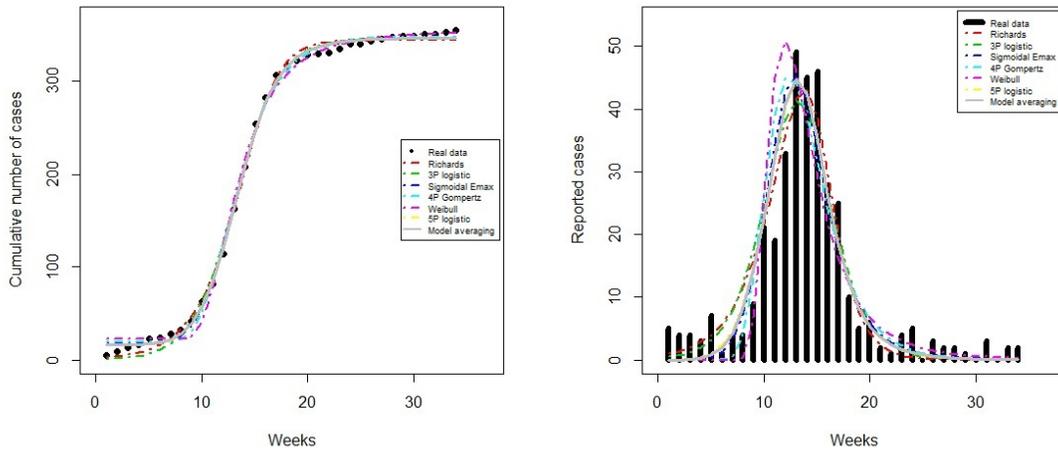
```
# Note que las opciones gráficas which=3 y which=4 no son utilizadas
# cuando se utiliza un solo modelo
> h1 <-allmodels(dengueoutbreak1$Incidence,dengueoutbreak1$Time, model="Richards")
> plot(h1,which=c(1,2), xlab="Weeks")
```

La figura 5.1, además de los datos del brote y los modelos ajustados, muestra la estimación puntual y por intervalo de confianza para el acmé y tamaño final de la epidemia, para todos los modelos incorporados al paquete de R.

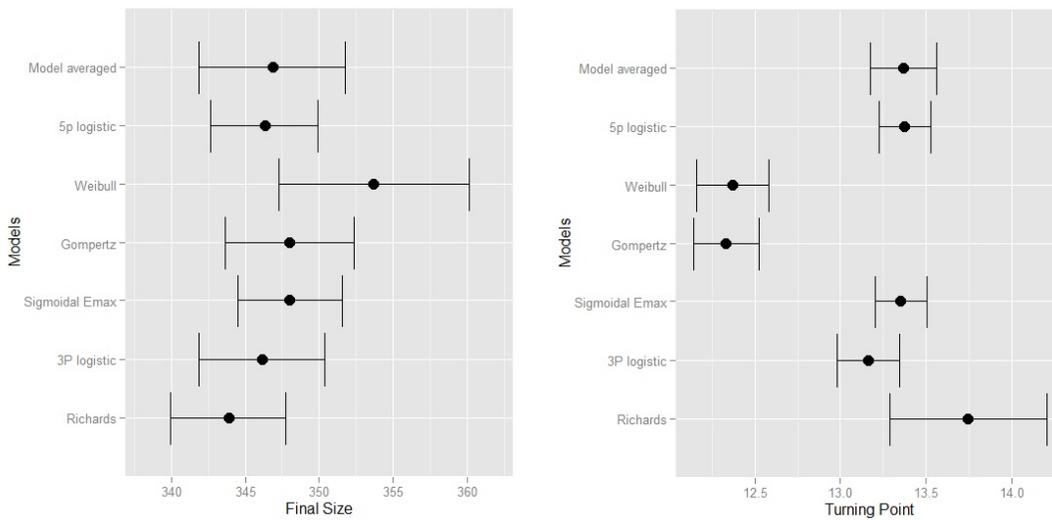
```
# Figure 5.1 a y 5.1 b
> h2 <-allmodels(dengueoutbreak2$Incidence,dengueoutbreak2$Time,model="all")
> plot(h2,which=c(1:4),xlab="Weeks")
```

5.2.2 Predicciones en tiempo real. Ejemplo de aplicación en dengue

La modelación basada en el promedio de modelos, o en un solo modelo, es particularmente útil para la predicción en tiempo real (es decir, se utiliza solo una parte de los datos para la predicción a largo plazo). Dada la incidencia por aparición de los síntomas o fecha de hospitalización durante el brote, se puede utilizar el paquete de R *DengueRT* [78] para



(a) Datos y modelos ajustados.



(b) Intervalo de confianza 95% para el tamaño final y acmé de la epidemia.

Figura 5.1: Ejemplo de salidas de la función `plot()` cuando se utilizan todos los modelos.

pronosticar la eventual severidad de las fases actuales del brote en tiempo real estimando la capacidad de carga. Sin embargo, la precisión depende en tener disponibles datos de incidencia para algunos tiempos posterior al acmé de la epidemia y bajo el supuesto que no habrá nuevas ondas de infección en el futuro. Las dos funciones principales del paquete *DengueRT* [78] para proporcionar predicciones en tiempo real son `allmodelpredict()` y `changetimeFSTP()`.

Suponga que están disponible los datos de incidencia del área de salud 1 hasta la semana 22 y se requiere obtener predicciones del tamaño final y del acmé de la epidemia al finalizar la epidemia, así como la incidencia y el número de casos acumulados en la semana 30.

```
# Predicción en tiempo real utilizado todos los modelos no lineales
# Datos de incidencia disponibles hasta la semana 22 (Área de salud 1)
> hrp2 <-allmodelpredict(dengueoutbreak1$Incidence[1:22],dengueoutbreak1$Time[1:22],30,
+ model = "all")
> summary(hrp2)
```

Los valores de AIC y los pesos de los modelos se muestran debajo.

```
#Salida 3
AIC
      Richards      3P Logistic Sigmoidal Emax      Gompertz      Weibull
117.6249      160.8052      169.1075      179.6142      185.8068
5P Logistic
129.3018
Model weights
      Richards      3P Logistic Sigmoidal Emax      Gompertz      Weibull
9.97095e-01      4.19025e-10      6.59831e-12      3.45087e-14      1.56036e-15
5P Logistic
2.90488e-03
```

Las predicciones específica de cada modelo y del promedio de modelo para tamaño final y acmé de la epidemia están dados por:

	lower	est.	upper
Richards	358.7022	363.4467	368.1911
3P logistic	373.0097	389.2167	405.4237

```

SigmEmax      375.4903 402.4293 429.3684
Gompertz      378.5680 419.4317 460.2954
Weibull       384.1957 472.9414 561.6870
5P logistic   355.9536 362.1653 368.3770
Model averaged 357.6308 363.4429 369.2550

```

```

          lower    est.    upper
Richards   15.79110 15.98481 16.17853
3P logistic 14.79243 15.08987 15.38731
SigmEmax   14.88398 15.31167 15.73936
Gompertz   13.77158 14.26487 14.75817
Weibull    13.75640 14.68116 15.60592
5P logistic 14.96065 15.06854 15.17643
Model averaged 15.73889 15.98215 16.22541

```

Model averaged prediction of the incidence at the time point 30

```
[1] 0.0003387103
```

Model averaged prediction of the cumulative number of cases at the time point 30

```
[1] 363.4426
```

Como con la función `allmodels()`, los resultados pueden ser visualizados utilizando la función `plot` con el mismo argumento de antes.

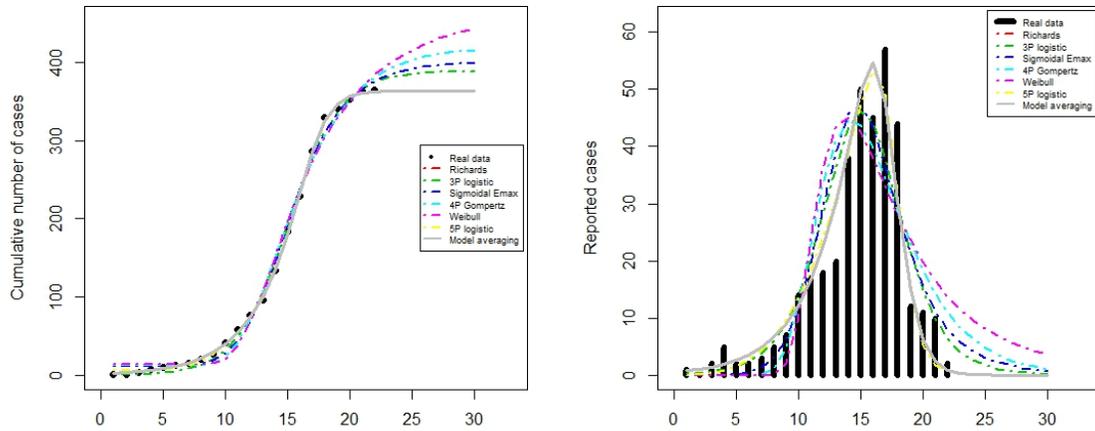
La figura 5.2, presenta los datos del brote hasta la semana 22 y los modelos ajustados hasta la semana 30 cuando son utilizados todos los modelos no lineales incorporados al paquete *DengueRT*.

```

# Figuras 5.2 a, 5.2 b
# Note que se utilizó solo las opciones de gráfico which=c(1,2)
> hrp2 <-allmodelpredict(dengueoutbreak1$Incidence[1:22],dengueoutbreak1$Time[1:22],
30, model = "all")
> plot(hrp2,which=c(1,2))

```

La última función principal es `changetimeFSTP()`. Esta función proporciona los cambios en el tiempo de las estimaciones del tamaño final y acmé de la epidemia (para cada modelo específico y para el promedio de modelos) desde el punto en el tiempo solicitado hasta el



(a) Modelos pronósticos casos acumulados.

(b) Modelos pronósticos de la incidencia.

Figura 5.2: Ejemplos de salidas de la función `plot()` cuando todos los modelos son utilizados para llevar a cabo predicciones en tiempo real.

último dato disponible. La salida de la función muestran dos tablas, una para el tamaño final y otra para el acmé de la epidemia.

Suponga que los datos de incidencia en el área de salud 2 están disponibles hasta la semana 34 y se quiere analizar el cambio en el tiempo del tamaño final y el acmé de la epidemia desde la semana 19 hasta el fin de la epidemia.

```
## (Área de salud 2)
## Cambios en el tiempo desde la semana 19 hasta la semana 34 de las
## estimaciones del tamaño final y el acmé de la epidemia para cada modelos
## y la estimación por promedio de modelos
> ct1 <-changetimeFSTP(dengueoutbreak2$Incidence,dengueoutbreak2$Time,ini=19)
> summary(ct1)
```

Los cambios en el tiempo de las estimaciones del acmé de la epidemia se muestran debajo (las filas representan los datos utilizados para la estimación del parámetro del modelo, es decir, en la primera fila fueron utilizadas las primeras 19 semanas, en la segunda fila las primeras 20 semanas, etc.).

Changes over time of the parameter estimates for the turning point

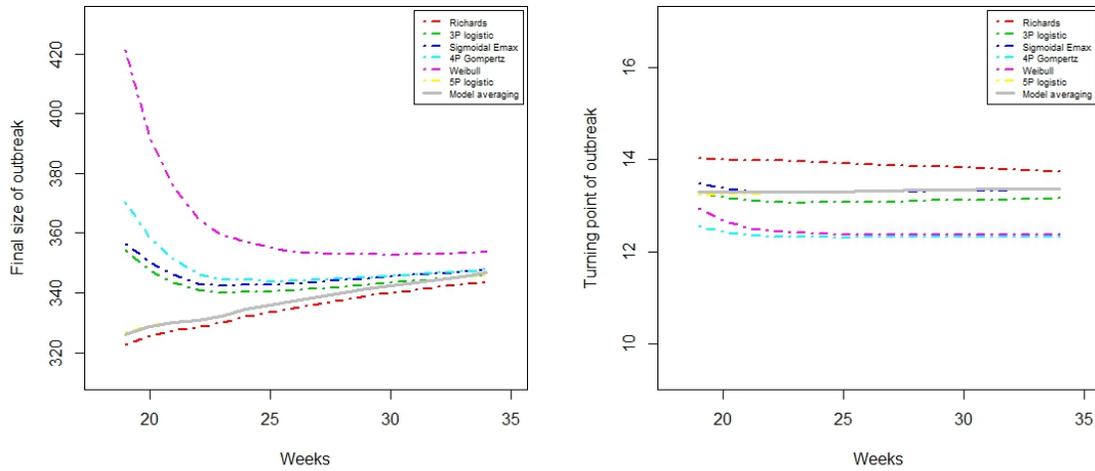
	Richards	logistic3P	SigmEmax	Gompertz	Weibull	logistic5P	Model averaged
1-19	14.02646	13.28732	13.46911	12.53888	12.94928	13.22999	13.31139
1-20	14.01093	13.18380	13.38643	12.43053	12.67944	13.25221	13.30232
1-21	13.99784	13.12017	13.33196	12.36903	12.53247	13.26291	13.29506
1-22	13.98775	13.07874	13.29457	12.33237	12.44737	13.26822	13.28977
1-23	13.97135	13.06464	13.28186	12.31760	12.40741	13.27937	13.29288
1-24	13.94743	13.06954	13.28617	12.31541	12.39158	13.29529	13.30502
1-25	13.92910	13.07124	13.28634	12.31251	12.37788	13.30523	13.31152
1-26	13.90808	13.07927	13.29210	12.31328	12.37164	13.31606	13.32045
1-27	13.88672	13.08937	13.29952	12.31529	12.36850	13.32618	13.32928
1-28	13.86482	13.10104	13.30822	12.31809	12.36740	13.33583	13.33797
1-29	13.84430	13.11203	13.31618	12.32071	12.36678	13.34412	13.34530
1-30	13.82680	13.12088	13.32214	12.32259	12.36587	13.35049	13.35075
1-31	13.80606	13.13285	13.33095	12.32564	12.36666	13.35802	13.35701
1-32	13.78789	13.14289	13.33802	12.32803	12.36702	13.36408	13.36165
1-33	13.76829	13.15415	13.34619	12.33083	12.36797	13.37047	13.36615
1-34	13.74728	13.16641	13.35525	12.33392	12.36931	13.37713	13.37064

Como antes, la función `plot()` puede ser utilizadas para la visualización de los resultados (con las opciones `which=5` para el tamaño final y `which=6` para el acmé de la epidemia).

```
# Figuras 5.3 a, 5.3 b
# Note que son utilizadas las opciones gráficas which=c(5,6)
# Datos de incidencia disponibles hasta la semana 34
# Se requiere analizar el cambio en el tiempo desde la semana 19
> ct1 <-changetimeFSTP(dengueoutbreak2$Incidence,dengueoutbreak2$Time,ini=19)
> plot(ct1,which=c(5,6))
```

5.3 Conclusiones parciales

Todos los modelos no lineales de crecimiento y el método de promedio de modelos presentados y discutidos en el capítulo 2 están implementados en el paquete de R, *DengueRT* [78], el cual está disponible en el *Comprehensive R Archive Network* (**CRAN**) en el sitio Web



(a) Cambios en el tiempo del tamaño final. (b) Cambios en el tiempo del acmé de la epidemia.

Figura 5.3: Ejemplos de salidas de la función `plot()` cuando el argumento es un objeto `changetimeFSTP` con las opciones de gráfico `which=c(5,6)`.

<https://cran.r-project.org/web/packages/DengueRT/>. En este capítulo se hace una introducción a este paquete, describiendo como se implementaron las funciones y como se creó el paquete desde el **RStudio** [74] y con el uso de la plataforma **R-forge** [82].

El paquete **DengueRT** [78] incluye las funciones principales `allmodels()`, `allmodelpredict()` y `changetimeFSTP()` para el cálculo de los resultados, la función `summary()` para producir resúmenes de los resultados y la función `plot()` para visualizar los resultados. Las principales características y las funcionalidades del paquete **DengueRT** han sido ilustradas utilizando dos conjuntos de datos correspondientes a un brote de una onda de dengue ocurrido en un municipio de La Habana en el 2001-2002.

Como las funciones implementadas en este paquete requieren solo de los datos de incidencia, el uso de este paquete puede ser extendido a otras enfermedades infecciosas tales como Zika, H1N1, SRAS, Ébola, etc. Este enfoque de modelación y su implementación en el paquete de R, **DengueRT** [78] pudiera ser una valiosa herramienta para los decisores de salud pública para enfrentar futuros brotes de enfermedades.

Conclusiones

- La metodología propuesta basada en el uso de promedio de modelos mostró su utilidad para la estimación y predicción en tiempo real de parámetros epidemiológicos en brotes de dengue de una onda y tiene la ventaja de que incorpora la incertidumbre implícita en la selección de modelos.
- La metodología diseñada para la modelación de brotes multiondas utilizando la verosimilitud perfil, a diferencia de la metodología clásica, permite estimar los parámetros epidemiológicos de cada onda simultáneamente.
- Para la comparación de brotes de dengue de una onda se proponen tres enfoques metodológicos que tienen en cuenta la incertidumbre de modelos utilizando el promedio de modelos Frecuentista y Bayesiano a través de los modelos jerárquicos y de la selección Gibbs de variables. Este último enfoque resulta más eficiente ya que permite estimar simultáneamente los pesos de todos los posibles modelos y los parámetros.
- La utilización de un modelo no lineal con efectos mixtos constituye una forma eficaz de tener en cuenta la heterogeneidad de la transmisión dentro de áreas en una misma región. El modelo no lineal individual para cada área puede sugerir el tipo de estructura de efectos aleatorios y los valores iniciales de los parámetros a utilizar en el modelo mixto.
- Se mostró la aplicabilidad de las metodologías propuestas en la estimación y predicción en tiempo real de parámetros epidemiológicos a los datos de brotes de dengue ocurridos en Cuba (2001-2002, 2006, 2007) y a un brote de Zika ocurrido en Colombia (2015-2016).
- El paquete *DengueRT* creado en R y disponible en el CRAN, permitió implementar computacionalmente la metodología diseñada para la estimación y predicción de parámetros epidemiológicos en brotes de dengue de una onda.

Recomendaciones

- Extender la metodología para la modelación de brotes de dengue multiondas al uso de otros modelos no lineales para realizar inferencias multi-modelos a través del promedio de modelo.
- Extender la metodología del uso del modelo no lineal con efectos mixtos para tener en cuenta la heterogeneidad entre las áreas al uso de otros modelos y a la modelación jerárquica bayesiana.
- Incorporar a los modelos propuestos para las áreas urbanas de una misma región los métodos de estimación para áreas pequeñas y métodos de imputación de datos faltantes.
- Incorporar nuevas funciones al paquete de R, *DengueRT* para la implementación las diferentes metodologías presentadas en esta investigación.
- Crear una interfaz gráfica del paquete para los epidemiológicos, de modo que facilite la aplicación de los modelos propuestos y la introducción en la práctica de los resultados de la investigación.

Referencias

- [1] AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (Budapest, 1973), B. N. Petrov and F. Csaki, Eds., Akadémiai Kiado, pp. 267–281.
- [2] AMINI, S. M. *Model uncertainty and model averaging techniques*. PhD thesis, Faculty of the Virginia Polytechnic Institute and State University, 2012.
- [3] ANDERSON, R. M., AND MAY, R. M. *Infectious diseases of humans*. Oxford University Press, Oxford, UK, 1991.
- [4] ANDO, T. *Bayesian Model Selection and Statistical Modeling*. A Chapman and Hall/CRC Press, 2010.
- [5] ANDRAUD, M., HENS, N., MARAIS, C., AND BEUTELS, P. Dynamic epidemiological models for dengue transmission: A systematic review of structural approaches. *PLoS ONE* 7, 11 (2012), e49085.
- [6] BATES, D., AND PINHEIRO, J. *Mixed Effects Models in S and Splus*. Statistics and Computing. Springer, New York, 2000.
- [7] BHATT, S., GETHING, P., BRADY, O., MESSINA, J., FARLOW, A., MOYES, C., DRAKE, J., BROWNSTEIN, J., HOEN, A., SANKOH, O., MYERS, M., GEORGE, D., JAENISCH, T., WINT, G., SIMMONS, C., SCOTT, T., FARRAR, J., AND HAY, S. The global distribution and burden of dengue. *Nature* 496 (2013).

- [8] BRAUER, F., AND CASTILLO-CHAVEZ, C. *Mathematical Models in Population Biology and Epidemiology*, second ed. Springer, New York, 2012.
- [9] BURNHAM, K., AND ANDERSON, D. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods Researchs* 33, 2 (November 2004), 261–304.
- [10] BURNHAM, K., AND ANDERSON, D. R. *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*, second ed. Springer-Verlag, New York, 2002.
- [11] CASELLA, G., AND GEORGE, E. Explaining the Gibbs sampler. *Journal of the American Statistical Association* 46 (1992), 167–174.
- [12] CASELLA, G., AND ROBERTS, C. *Monte Carlo Statistical Methods*, 2nd edition ed. Springer-Verlag, New York, 2004.
- [13] CHOWELL, G., FUENTES, R., OLEA, A., AGUILERA, X., NESSE, H., AND HYMAN, J. The basic reproduction number R_0 and effectiveness of reactive interventions during dengue epidemics: The 2002 dengue outbreak in easter island, Chile. *Math Biosci Eng.* 10 (2013), 1455–1474.
- [14] CHOWELL, G., HINCAPIE-PALACIO, D., OSPINA, J., PELL, B., TARIQ, A., DAHAL, S., MOGHADAS, S., SMIRNOVA, A., SIMONSEN, L., AND VIBOUD, C. Using phenomenological models to characterize transmissibility and forecast patterns and final burden of Zika epidemics. *PLOS Currents Outbreaks* (May 2016).
- [15] CHOWELL, G., SATTENSPIEL, L., BANSALD, S., AND VIBOUD, C. Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews* (July 2016).
- [16] CHOWELL, G., AND VIBOUD, C. Is it growing exponentially fast? -Impact of assuming exponential growth for characterizing and forecasting epidemics with initial near-exponential growth dynamics. *Infectious Disease Modelling* (2016).
- [17] CHOWELL, G., VIBOUD, C., SIMONSEN, L., AND MOGHADAS, S. M. Characterizing the reproduction number of epidemics with early subexponential growth dynamics. *J. R. Soc. Interface* 13, 20160659 (2016).

- [18] CLAESKENS, G., AND HJORT, N. L. *Model selection and model averaging*. Cambridge University Press, 2008.
- [19] COOKE, R. *Uncertainty modeling in dose response: bench testing environmental toxicity*. Statistics in Practice. John Wiley & Sons, Inc., Hoboken, New Jersey, 2009.
- [20] COTTON, R. *Learning R*. O Reilly Media, Inc., Sebastopol, CA, 2013.
- [21] CRAWLEY, M. J. *Statistics : an introduction using R*, second ed. John Wiley & Sons, Ltd, United Kingdom, 2015.
- [22] DE VRIES, A., AND MEYS, J. *R For Dummies*, 2nd edition ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2015.
- [23] DEMIDENKO, E. *Mixed models: theory and applications with R*, second edition ed. Wiley series in probability and statistics. John Wiley & Sons, 2013.
- [24] EFRON, B., AND TIBSHIRANI, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1, 1 (1986), 54–75.
- [25] ELOUSA, P. *Introducción al entorno R*. Marcel Dekker Inc, Ed. Universidad del País Vasco, 2011.
- [26] FAES, C., AERTS, M., GEYS, H., AND MOLENBERGHS, G. Model averaging using fractional polynomials to estimate a safe level of exposure. *Risk Analysis* 27, 1 (February 2007), 111–123.
- [27] FIRAT, M., KARAMAN, E., BASAR, E., AND NARINC, D. Bayesian analysis for the comparison of nonlinear regression model parameters: an application to the growth of japanese quail. *Brazilian Journal of Poultry Sciense Special issue, Quails* (2015), 19–26.
- [28] GELFAND, A., KIM, H., SIRMANS, C., AND BANERJEE, S. Spatial modeling with spatially varying coeficient processes. *Journal of the American Statistical Association* 98 (2003), 387–396.

- [29] GELFAND, A., AND SMITH, A. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85 (1990), 398–409.
- [30] GILKS, W. R., RICHARDSON, S., AND SPIEGELHALTER, D. J. *Markov Chain Monte Carlo in practice*, reprint ed ed. Chapman and Hall, London, 1996.
- [31] GUBLER, D. J., OOI, E. E., VASUDEVAN, S., AND FARRAR, J., Eds. *Dengue and Dengue Hemorrhagic Fever*, 2nd edition ed. C.A.B. International, Boston, USA, 2014.
- [32] GUZMÁN, M. G., AND OTROS. *Dengue*. Editorial Ciencias Médicas, La Habana, 2014.
- [33] GUZMÁN, M. G., PELÁEZ, O., KOURÍ, G., QUINTANA, I., VÁZQUEZ, S., PENTÓN, M., AND ÁVILA, L. C. Caracterización final y lecciones de la epidemia de dengue 3 en Cuba, 2001-2002. *Revista Panamericana de Salud Pública* 19, 4 (2006), 282–289.
- [34] HAMRA, G., MACLEHOSE, R., AND RICHARDSON, D. Markov Chain Monte Carlo: An introduction for epidemiologists. *International Journal of Epidemiology* 42 (2013), 627–634.
- [35] HSIEH, Y. Richards model: A simple procedure for real-time prediction of outbreak severity. In *Modeling and Dynamics of Infectious Diseases (Volumen 11)*, Z. Ma, J. Wu, and Y. Zhoue, Eds., Contemporary Applied Mathematics (CAM). Higher Education Press, 2009, pp. 216–236.
- [36] HSIEH, Y. Temporal course of 2014 Ebola virus disease (EVD) outbreak in west africa elucidated through morbidity and mortality data: a tale of three countries. *PLoS ONE* 10, 11 (2015), e0140810.
- [37] HSIEH, Y. Temporal patterns and geographic heterogeneity of Zika virus (ZIKV) outbreaks in french polynesia and central america. *PeerJ*, 5 (2017), e3015.
- [38] HSIEH, Y., ARAZOZA, H., AND LOUNES, R. Temporal trends and regional variability of 2001-2002 multiwave DENV-3 epidemic in Havana City: did hurricane Michelle contribute to its severity? *Tropical Medicine and International Health* 18, 7 (July 2013), 830–838.

- [39] HSIEH, Y., AND CHEN, C. Turning points, reproduction number, and impact of climatological events for multi-wave dengue outbreaks. *Tropical Medicine and International Health* 14, 6 (January 2009), 628–638.
- [40] HSIEH, Y., AND CHENG, Y. Real-time forecast of multi-wave epidemic outbreaks. *Emerging Infectious Diseases* 12, 1 (January 2006), 122–127.
- [41] HSIEH, Y., HUANG, H., AND LAN, Y. On temporal patterns and circulation of influenza virus strains in Taiwan, 2008-2014: Implications of 2009 pH1N1 pandemic. *PLoS ONE* 11, 5 (2016), e0154695.
- [42] HSIEH, Y., AND MA, S. Intervention measures, turning point, and reproduction number for dengue, Singapore, 2005. *American Journal of Tropical Medicine and Hygiene* 80, 1 (January 2009), 66–71.
- [43] HSIEH, Y., MA, S., VELASCO, J., LEE, V., AND LIM, W. Early outbreak of 2009 influenza A (H1N1) in Mexico prior to identification of pH1N1 virus. *PLoS ONE* 6, 8 (2011), e23853.
- [44] JOHANSSON, M., HOMBACH, J., AND CUMMINGS, D. Models of the impact of dengue vaccines: A review of current research and potential approaches. *Vaccine* 29 (2011), 5860–5868.
- [45] KIM, S., AND SANDERS, N. Model averaging with AIC weights for hypothesis testing of hormesis at low doses. *Dose-Response* 15, 2 (2017), 1559325817715314.
- [46] LEISCH, F. Creating R packages: A tutorial. In *Compstat 2008 - Proceedings in Computational Statistics*, P. Brito, Ed. Physica-Verlag, Heidelberg, Germany, 2008.
- [47] LESAFFRE, E., AND LAWSON, A. *Bayesian Biostatistics*. John Wiley & Sons, Ltd., United Kingdom, 2012.
- [48] LI, J., AND LOU, Y. Characteristics of an epidemic outbreak with a large initial infection size. *Journal of Biological Dynamics* 10, 1 (2016).

- [49] LIANG, F., LIU, C., AND CARROLL, R. *Advanced Markov Chain Monte Carlo Methods*. Wiley, 2010.
- [50] LIAO, J., AND LIU, R. Re-parameterization of five-parameter logistic function. *Journal of Chemometrics* 23, 5 (May 2009), 248–253.
- [51] LIN, D., SHKEDY, Z., YEKUTIELI, D., AMARATUNGA, D., AND BIJNENS, L. *Modeling Dose-response Microarray Data in Early Drug Development Experiments Using R, Order Restricted Analysis of Microarray Data*, vol. 63 of *Use R !* Springer-Verlag Berlin Heidelberg, 2012.
- [52] LINDSTROM, M., AND BATES, D. Nonlinear mixed-effects models for repeated measures data. *Biometrics* 46 (1990), 673–687.
- [53] LUND, D., AND SORENSEN, K. H. Credit scoring a Bayesian approach. Master’s thesis, Aarhus University, Business and social science, 2012.
- [54] MA, J., DUSHOFF, J., BOLKER, B., AND EARN, D. Estimating initial epidemic growth rate. *Bull Math Biol* 76 (2014), 245–260.
- [55] MACDOUGALL, J. Analysis of dose-response studies-Emax model. In *Dose Finding in Drug Development*, N. Ting, Ed., Statistics for Biology and Health. Springer New York, 2006, pp. 127–145.
- [56] MASSAD, E., AND COUTINHO, F. A. Estimation of R_0 from the initial phase of an outbreak of a vector-borne infection. *Tropical Medicine and International Health* 15, 1 (2010), 120–126.
- [57] MOON, H., KIM, H., CHEN, J., AND KODELL, R. Model averaging using the kullback information criterion in estimating effective doses for microbial infection and illness. *Risk Analysis* 25, 5 (October 2005), 1147–1159.
- [58] NGUEFACK-TSAGUE, G. On optimal weighting scheme in model averaging. *American Journal of Applied Mathematics and Statistics* 2, 3 (2014), 150–156.

- [59] NISHIURA, H. Mathematical and statistical analyses of the spread of dengue. *Dengue Bulletin* 30 (2006), 51–67.
- [60] NTZOUFRAS, I. *Bayesian Modeling Using WinBUGS*. John Wiley & Sons Inc., Hoboken, New Jersey, 2009.
- [61] PACE, L. *Beginning R: An Introduction to Statistical Programming*. Springer, New York, 2012.
- [62] PELÁEZ, O., GUZMÁN, M. G., KOURÍ, G., PÉREZ, R., MARTÍN, J. L. S., VÁZQUEZ, S., ROSARIO, D., MORA, R., QUINTANA, I., BISSET, J., CANCIO, R., MASA, A. M., CASTRO, O., GONZÁLEZ, D., ÁVILA, L. C., RODRÍGUEZ, R., ÁLVAREZ, M., PELEGRINO, J. L., BERNARDO, L., AND PRADO, I. Dengue 3 epidemic, Havana, 2001. *Emerging Infectious Disease* 10, 4 (2004), 719–722.
- [63] PENG, R. D. *R Programming for Data Science*. Leanpub, 2015.
- [64] PINHEIRO, J., BATES, D., DEBROY, S., SARKAR, D., AND R CORE TEAM. *nlme: Linear and Nonlinear Mixed Effects Models*, 2016. R package version 3.1-128.
- [65] POSADA, D., AND BUCKEY, T. R. Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology* 53, 5 (2004), 793–808.
- [66] R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.
- [67] R CORE TEAM. *Writing R extensions*. R Foundation for Statistical Computing, Vienna, Austria, 2016.
- [68] RACLOZ, V., RAMSEY, R., TONG, S., AND HU, W. Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. *Public Library of Science Neglected Tropical Diseases* 6, 5 (2012), e1648.
- [69] REINER, R., PERKINS, T., BARKER, C., NIU, T., CHAVES, L., ELLIS, A., GEORGE, D., MENACH, A. L., PULLIAM, J., BISANZIO, D., CHIAYAKA, C., CUMMINGS, D.,

- GARCIA, A., GATTON, M., GETHING, P., HARTLEY, D., JOHNSTON, G., KLEIN, E., MICHAEL, E., LINDSAY, S., LLOYD, A., PIGOTT, D., REISEN, W., RUKTANONCHAI, N., SINGH, B., TATEM, A., KITRON, U., HAY, S., SCOTT, T., AND SMITH, D. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970-2010. *Journal of the Royal Society Interface* 10, 81 (2013), 20120921.
- [70] RICHARDS, F. A flexible growth function for empirical use. *Journal of Experimental Botany* 10, 29 (1959), 290–301.
- [71] RITZ, C., AND STREIBIG, J. C. Bioassay analysis using R. *Journal of Statistical Software* 12 (2005).
- [72] ROBERT, C. P. The Metropolis-Hastings algorithm. *Wiley StatsRef: Statistics Reference Online* (2016).
- [73] ROBERT, C. P., AND CASELLA, G. *Introducing Monte Carlo Method with R*. Springer, Berlin, 2010.
- [74] RSTUDIO TEAM. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2016.
- [75] SÁNCHEZ, L., VANLERBERGHE, V., ALFONSO, L., MARQUETTI, M., GUZMÁN, M., BISSET, J., AND DER STUYFT, P. V. Aedes aegypti larval indices identify neighbourhood high risk for dengue epidemics. *Emerging Infectious Disease* 12 (2006), 800–806.
- [76] SAS INSTITUTE INC. *SAS/STAT Software, Version 9.4*. Cary, NC, 2013.
- [77] SEBER, G., AND WILD, C. *Nonlinear regression*. John Wiley & Sons, New Jersey, 2003.
- [78] SEBRANGO, C., SANCHEZ, L., SHKEDY, Z., AND DE TROYER, E. *DengueRT: Parameter Estimates and Real-Time Prediction of a Single Dengue Outbreak*, 2016. R package version 1.0.1.
- [79] SMIRNOVA, A., AND CHOWELL, G. A primer on stable parameter estimation and forecasting in epidemiology by a problem-oriented regularized least squares algorithm. *Infectious Disease Modelling* 2, 2 (2017), 268–275.

- [80] SMITH, A., AND ROBERTS, G. Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society B* 55, 1 (1993), 3–23.
- [81] SMITH, D., BATTLE, K., HAY, S., BARKER, C., SCOTT, T., AND MCKENZIE, F. Ross, Macdonald and a theory for the dynamics and control of mosquito-transmitted pathogens. *Public Library of Science Pathogens* 8 (2012), e1002588.
- [82] THEUßL, S., AND ZEILEIS, A. Collaborative Software Development Using R-Forge. *The R Journal* 1, 1 (2009), 9–14.
- [83] TSOULARIS, A., AND WALLACE, J. Analysis of logistic growth models. *Mathematical Biosciences* 179, 1 (July-August 2002), 21–55.
- [84] WALLINGA, J., AND LIPSITCH, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of Royal Society Biological Sciences* 274 (February 2007), 599–604.
- [85] WANG, X., WU, J., AND YANG, Y. Richards model revisited: validation by and application to infection dynamics. *Journal of Theoretical Biology* 313 (2012), 12–19.
- [86] WICKHAM, H. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.
- [87] WICKHAM, H. *Advanced R*. The R series. CRC Press, Taylor & Francis Group, Boca Raton, FL, 2015.

ANEXOS

Anexo 1

Producción científica del autor acerca del tema de la tesis.

Publicaciones.

Sebrango C., Sanchez L., Shkedy Z. (2017) Modeling dengue outbreak data using nonlinear mixed effects model: Revista Investigación Operacional Vol. 38 No. 3, 252-258.

Sebrango-Rodriguez C.R., Martinez-Bello D.A., Sanchez-Valdes L., Thilakarathne P. J., Del Fava E., Van der Stuij P., López-Quílez A., Shkedy Z. (2017) Real-time parameter estimation of Zika outbreaks using model averaging: Epidemiology and Infection, Vol. 145 No. 11, 2313-2323, doi: 10.1017/S0950268817001078

Sebrango C., Sanchez L., Shkedy Z., De Troyer E., DengueRT: Parameter Estimates and Real-Time Prediction of a Single Dengue Outbreak, R package version 1.0.1 (2016). URL <https://CRAN.R-project.org/package=DengueRT>

Sanchez L., **Sebrango C.**, Vega B., Alfonso L. (2014) Capítulo 3: Modelos de transmisión, pronóstico y gravedad del dengue. En Guzman M. G. y otros, Dengue. La Habana: Editorial Ciencias Médicas.

Sebrango C., Carbonell N., Sanchez L., Shkedy Z. (2012) Estimación de parámetros de transmisión para un brote epidémico de dengue en La Habana: Revista Ciencias Matemáticas Vol. 26 No. 1, 31-36.

Sanchez L., **Sebrango C.**, Abad Y., Pelaez O., Shkedy Z. y Van der Stuyft P. (2011) Reproduction number, turning point, and total number of cases during a dengue outbreak in Playa Municipality, Havana City. Tropical Medicine and International Health; 16 (Suppl. 1): 243-244.

Sanchez L., Vega B., **Sebrango C.**, Carbonell N., Cristo M., Rodriguez C., Pérez K., Alfonso L., Castro M., Rodriguez-Roche R., Marquetti M.C., Cruz C., Rios L., González D., Castro O., Toledo I., Concepción D. (2011) Contribuciones de la modelación de enfermedades infecciosas en Cuba. Parte I: Aplicación al análisis epidemiológico del dengue. Revista Boletín Epidemiológico Semanal BOLIPK Vol. 21 No. 34: 265-271, Dirección Nacional de Epidemiología, Ministerio de Salud Pública ISSN 1028-5083.

Otras Publicaciones relacionadas con el tema.

Castro M, Sanchez L, Perez D, **Sebrango C**, Shkedy Z, y Van der Stuyft P. (2013) The Relationship between Economic Status, Knowledge on Dengue, Risk Perceptions and Practices. PLoS ONE 8(12): e81875. doi:10.1371/journal.pone.0081875

Cruz C., **Sebrango C.**, Cristo M., Pina C., Marquetti M.C.; Sanchez L. (2010) Comportamiento estacional y temporal de *Aedes aegypti* (Diptera: Culicidae) en Sancti Spíritus, 1999-2007. Revista Cubana Medicina Tropical; Vol. 62 No.1: 1-10.

Castro M., Gólvez C., Sanchez L., Pérez D., Polo V., Concepción D., **Sebrango C.** y Patrick Van der Stuyft P. (2010) Encuesta poblacional sobre conocimiento y percepciones acerca de dengue contra prácticas preventivas en el municipio La Lisa. Revista Cubana de Medicina Tropical. Vol.62 No.3: 245-253.

Premios.

Premio Nacional en la categoría de libro en el XLII CONCURSO "PREMIO ANUAL DE LA SALUD 2017", Libro: Dengue.

Premio Nacional de la Academia de Ciencias 2013: Nuevas evidencias sobre estrategias de empoderamiento comunitario para la prevención del dengue: Investigaciones de intro-

ducción a la práctica.

Premio Nacional de la Academia de Ciencias 2012: Contribuciones de la modelación matemática al análisis epidemiológico del dengue.

Participación en eventos como autor.

Parameter estimation and real-time prediction of a dengue outbreak using model averaging: The DengueRT R package. **XIII International Conference of Operational Research ICOR 2018**, Marzo 2018, La Habana.

La estimación de parámetros y predicción en tiempo real de un brote de dengue usando promedio de modelos: el paquete de R, DengueRT. **XV Congreso Internacional COMPUMAT 2017**, Noviembre 2017, La Habana.

El paquete de R, DengueRT: una contribución al control del dengue desde la Universidad. **Universidad 2018 Evento provincial**, junio 2017, Sancti Spiritus.

Modeling dengue outbreak data using nonlinear mixed effects model, **XII International Conference of Operational Research (ICOR)**, Marzo 2016, La Habana.

Modelación de datos de brotes de dengue y el impacto de intervenciones sociales. **Taller de Modelación matemática para el estudio de epidemias**, Septiembre 2016, Instituto de Medicina Tropical IPK, La Habana.

La estimación de parámetros y predicción en tiempo real de brotes de dengue de una onda usando el model averaging: El paquete de R DengueParmEst. **3. Conferencia Científica Internacional YAYABOCIENCIA 2015**, Noviembre 2015, Sancti Spiritus.

Estimación de parámetros y predicción en tiempo real de brotes de dengue de una onda us-

ando model averaging. **XIV Congreso Internacional COMPUMAT 2015**, Noviembre 2015, La Habana.

La estimación de parámetros y predicciones en tiempo real de brotes de dengue usando model averaging, su implementación en un paquete de R. **XVI Forum de Ciencia y Técnica Municipal**, Noviembre 2015, Sancti Spiritus.

Parameter estimation and real-time predictions of a dengue outbreak taking into account model uncertainty using model averaging (Poster), **Evento Internacional SIMID 2015**, Abril 2015, Amberes, Bélgica.

Estimación de parámetros y predicción en tiempo real de un brote epidémico de dengue usando model averaging. **2. Conferencia Científica Internacional YAYABOCIENCIA 2013**, Noviembre 2013, Sancti Spiritus.

Turning points, reproduction number and impact of intervention for Havana City 2001-2002 multi-wave dengue outbreak (Poster). **Evento Internacional SIMID 2012**, Abril 2012, Hasselt, Bélgica.

Turning points, reproduction number, and impact of intervention for Havana City 2001-2002 multi-wave dengue outbreak. (Poster). **Proceeding of the EPIDEMICS3- 3rd International Conference on Infectious Disease Dynamics**. Noviembre 2011, Boston, E.U.

Punto de retorno, número reproductivo básico e impacto de las intervenciones para un brote epidémico multiondas en La Ciudad de La Habana. **I Conferencia Internacional YAYABOCIENCIA 2011**, Noviembre 2011, Sancti Spiritus.

Estimación de parámetros de transmisión para un brote epidémico de dengue en la Habana. **XII Congreso Internacional COMPUMAT 2011**, Noviembre 2011, Villa Clara.

Anexo 2

Datos utilizados en la modelación

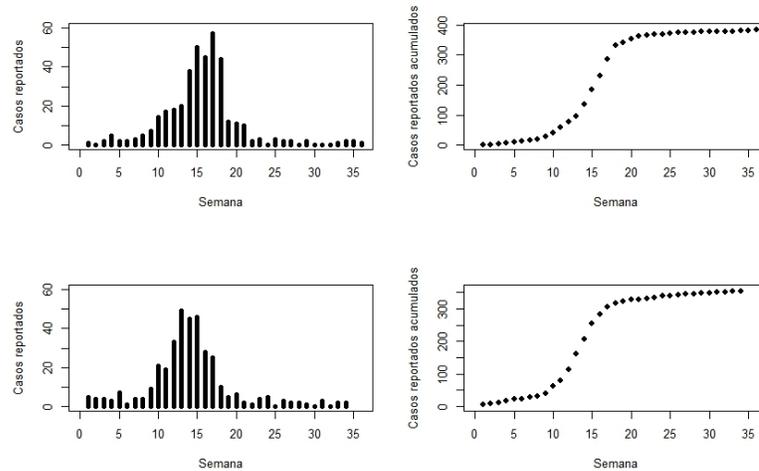


Figura 5.4: Incidencia y casos acumulados para el brote de dengue 2001/2002 en las áreas "26 de julio" (paneles superiores) y "J.R. Ramírez" (paneles inferiores), La Habana, Cuba

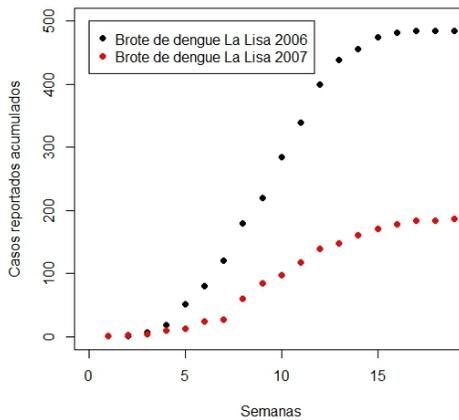


Figura 5.5: Casos reportados acumulados para los brotes de dengue 2006 y 2007 en el municipio La Lisa, La Habana, Cuba

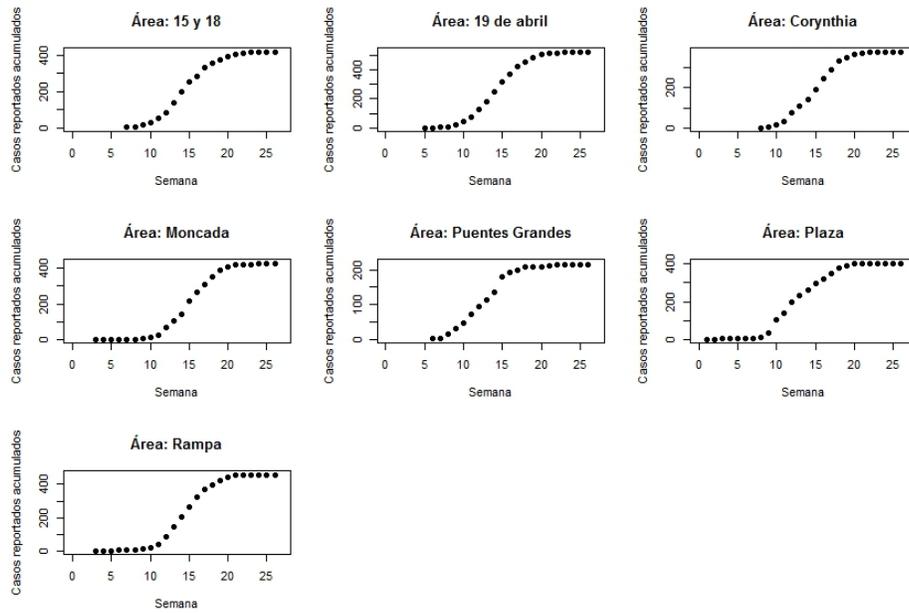


Figura 5.6: Número acumulado de casos observados en las siete áreas de salud del municipio Plaza, La Habana, afectados por el brote de dengue del 2006.

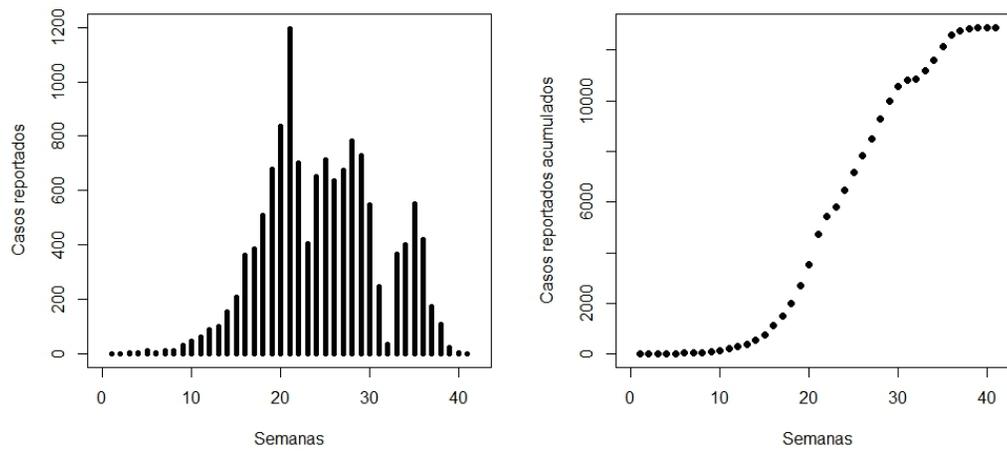


Figura 5.7: Incidencia y casos acumulados para el brote de dengue 2001/2002 en La Habana.

Anexo 3

Salidas del ejemplo de aplicación del Capítulo 2

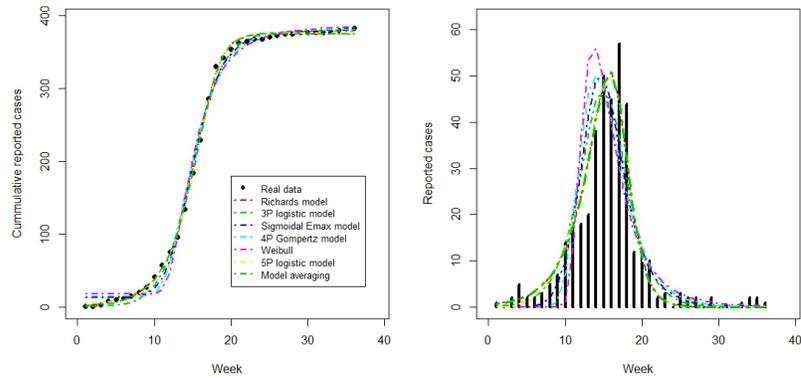


Figura 5.8: Área "26 de Julio". Panel de la izquierda: número acumulado de casos y los modelos ajustados. Panel de la derecha: datos de incidencia y los modelos ajustados.

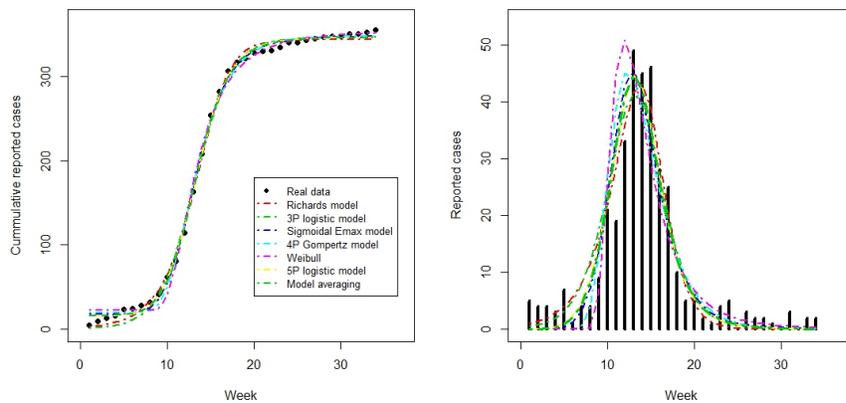


Figura 5.9: Área "J. R. Ramírez". Panel de la izquierda: número acumulado de casos y los modelos ajustados. Panel de la derecha: daos de incidencia y los modelos ajustados.

Tabla 5.2: Acné de la epidemia predicho usando todos los modelos y el promedio de modelos para diferentes períodos de tiempo en el área “26 de julio”.

Período	Logístico 3P	Richards	Logístico 5P	Sigmoideo Emax	Gompertz	Weibull	Promedio de modelos
1-18	17.18 (16.43,17.93)	16.18 (15.90,16.46)	14.98 (14.33,15.63)	21.12 (16.63,25.60)	22.99 (17.18,28.80)	1585.17 (0*,14720.58)	16.18 (15.80,16.57)
1-19	15.91 (15.28,16.54)	16.05 (15.85,16.26)	14.91 (14.78,15.05)	16.84 (15.32,18.36)	16.63 (14.38,18.88)	25.08 (8.11,42.00)	16.00 (15.59,16.41)
1-20	15.42 (14.95,15.88)	16.04 (15.85,16.24)	14.98 (14.87,15.09)	15.85 (15.02,16.69)	15.06 (13.96,16.16)	17.03 (13.30,20.76)	16.03 (15.77,16.29)
1-25	14.94 (14.73,15.16)	15.94 (15.75,16.13)	15.10 (15.00,15.21)	15.10 (14.81,15.38)	14.03 (13.69,14.37)	14.16 (13.73,14.58)	15.94 (15.69,16.18)
1-30	14.92 (14.75,15.09)	15.86 (15.64,16.08)	15.13 (15.02,15.24)	15.04 (14.83,15.26)	13.99 (13.70,14.27)	14.06 (13.74,14.38)	15.86 (15.57,16.15)
1-35	14.93 (14.78,15.07)	15.81 (15.58,16.06)	15.15 (15.03,15.26)	15.04 (14.85,15.23)	13.98 (13.73,14.24)	14.05 (13.76,14.34)	15.80 (15.47,16.13)
1-36	14.93 (14.79,15.08)	15.80 (15.55,16.05)	15.15 (15.03,15.27)	15.04 (14.85,15.23)	13.98 (13.73,14.23)	14.05 (13.77,14.33)	15.79 (15.43,16.14)

Tabla 5.3: Tamaño final predicho usando todos los modelos y el promedio de modelos para diferentes períodos de tiempo en el área “26 de julio”.

Período	Logístico 3P	Richards	Logístico 5P	Sigmoideo Emax	Gompertz	Weibull	Promedio de modelos
1-18	580.76 (501.20,660.31)	394.33 (339.86,448.80)	350.80 (287.05,414.56)	1019.44 (407.04,1631.83)	1821.19 (425.17,3217.20)	6444284.71 (0*,7005641)	396.69 (315.65,477.73)
1-19	453.09 (404.38,501.81)	353.95 (341.42,366.49)	344.30 (333.76,354.85)	529.14 (402.88,655.40)	682.33 (423.05,941.62)	1651.20 (0*,3734.85)	353.46 (337.70,369.22)
1-20	413.13 (382.43,443.84)	356.01 (348.77,363.25)	351.62 (343.66,359.58)	445.54 (384.51,506.58)	504.49 (396.47,612.52)	717.27 (346.28,1088.27)	355.96 (347.10,364.81)
1-25	379.04 (369.73,388.36)	367.23 (364.06,370.39)	366.95 (362.80,371.10)	385.05 (371.25,398.86)	389.77 (370.81,408.73)	408.05 (374.96,441.13)	367.22 (363.31,371.14)
1-30	377.39 (371.91,382.87)	371.85 (369.23,374.46)	372.00 (368.84,375.16)	380.55 (373.03,388.08)	381.89 (372.00,391.77)	389.71 (374.49,404.94)	371.85 (368.58,375.12)
1-35	377.93 (373.95,381.92)	374.47 (372.16,376.79)	374.73 (372.09,377.38)	380.09 (374.86,385.31)	380.72 (373.92,387.52)	385.52 (375.69,395.35)	374.48 (371.55,377.41)
1-36	378.29 (374.47,382.10)	375.08 (372.75,377.42)	375.37 (372.73,378.01)	380.35 (375.40,385.29)	380.91 (374.49,387.33)	385.42 (376.22,394.61)	375.09 (372.13,378.05)

Tabla 5.4: Acmé de la epidemia predicho usando todos los modelos y el promedio de modelos para diferentes períodos de tiempo en el área “J. R. Ramírez”.

Período	Logístico 3P	Richards	Logístico 5P	Sigmoideo Emax	Gompertz	Weibull	Promedio de modelos
1-16	14.82 (13.07,16.57)	14.00 (13.54,14.46)	12.92 (12.51,13.33)	14.55 (13.00,16.10)	14.55 (12.17,16.94)	21.48 (5.44,37.53)	13.28 (11.90,14.65)
1-17	13.99 (13.05,14.92)	14.03 (13.62,14.44)	13.19 (12.84,13.54)	14.02 (13.22,14.83)	13.38 (12.32,14.44)	15.46 (11.69,19.23)	13.36 (12.50,14.22)
1-18	13.53 (12.92,14.13)	14.03 (13.64,14.42)	13.22 (13.01,13.44)	13.67 (13.18,14.16)	12.81 (12.22,13.39)	13.64 (12.14,15.13)	13.33 (12.75,13.92)
1-20	13.18 (12.82,13.54)	14.01 (13.65,14.37)	13.25 (13.11,13.39)	13.39 (13.11,13.67)	12.43 (12.11,12.75)	12.68 (12.14,13.22)	13.30 (12.98,13.62)
1-25	13.07 (12.85,13.29)	13.93 (13.57,14.29)	13.31 (13.18,13.43)	13.29 (13.11,13.46)	12.31 (12.09,12.54)	12.38 (12.10,12.65)	13.31 (13.15,13.48)
1-30	13.12 (12.94,13.31)	13.83 (13.42,14.24)	13.35 (13.21,13.49)	13.32 (13.17,13.47)	12.32 (12.13,12.52)	12.37 (12.14,12.59)	13.35 (13.18,13.52)
1-34	13.17 (12.99,13.35)	13.75 (13.29,14.20)	13.38 (13.23,13.53)	13.36 (13.21,13.50)	12.33 (12.14,12.52)	12.37 (12.16,12.58)	13.37 (13.18,13.56)

Tabla 5.5: Tamaño final predicho usando todos los modelos y el promedio de modelos para diferentes períodos de tiempo en el área “J. R. Ramírez”.

Período	Logístico 3P	Richards	Logístico 5P	Sigmoideo Emax	Gompertz	Weibull	Promedio de modelos
1-16	475.29 (322.11,625.48)	290.66 (261.86,319.45)	296.51 (255.14,337.87)	442.70 (315.90,569.50)	590.07 (314.02,866.12)	1383.92 (0*,3385.42)	298.21 (243.63,352.79)
1-17	403.80 (338.79,468.81)	313.53 (289.55,337.52)	322.39 (287.74,357.04)	399.00 (340.06,457.94)	457.61 (352.98,562.24)	677.28 (298.96,1055.60)	322.20 (278.90,365.51)
1-18	370.39 (334.09,406.70)	319.93 (304.74,335.12)	325.95 (306.35,345.55)	371.28 (337.99,404.57)	398.12 (344.45,451.78)	491.01 (349.00,633.02)	325.29 (301.64,348.95)
1-20	347.52 (329.82,365.22)	325.54 (317.26,333.82)	329.04 (319.78,338.29)	350.16 (334.02,366.29)	358.46 (334.89,382.03)	392.28 (343.26,441.30)	328.81 (317.49, 340.13)
1-25	340.41 (332.91,347.91)	333.52 (328.84,338.20)	335.77 (331.26,340.29)	342.61 (336.11,349.11)	343.85 (335.10,352.60)	354.89 (339.66,370.12)	335.75 (330.16,341.35)
1-30	343.40 (338.35,348.45)	339.92 (335.80,344.04)	342.24 (338.40,346.09)	345.40 (341.17,349.63)	345.66 (340.23,351.09)	352.66 (344.03,361.28)	342.33 (337.38,347.28)
1-34	346.07 (341.79,350.35)	343.83 (339.90,347.76)	346.27 (342.63,349.92)	348.00 (344.60,351.58)	347.97 (343.60,352.35)	353.66 (347.23,360.08)	346.80 (341.81,351.78)

Anexo 4

Salidas del procedimiento a los datos de Zika en Colombia

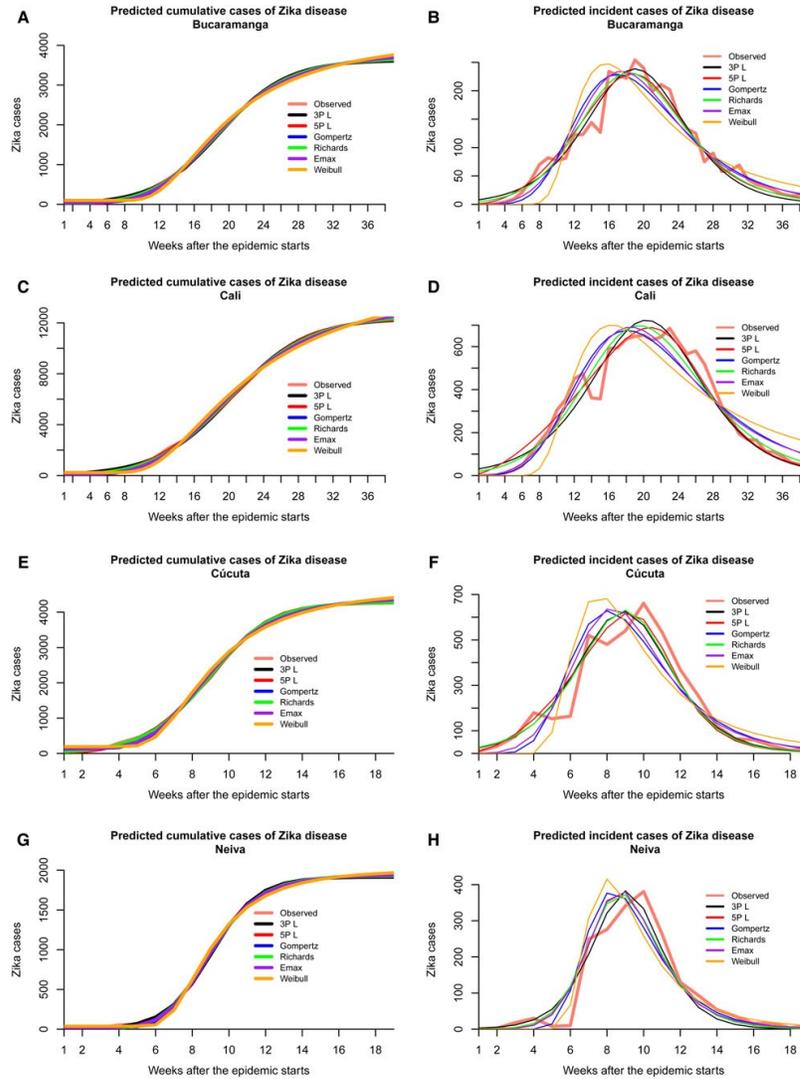


Figura 5.10: Predicción de los casos de incidencias y acumulados basados en seis modelos no lineales para el brote de Zika en cuatro ciudades colombianas, La predicción es hecha cuando se utilizan todos los datos para la estimación de los parámetros del modelo.

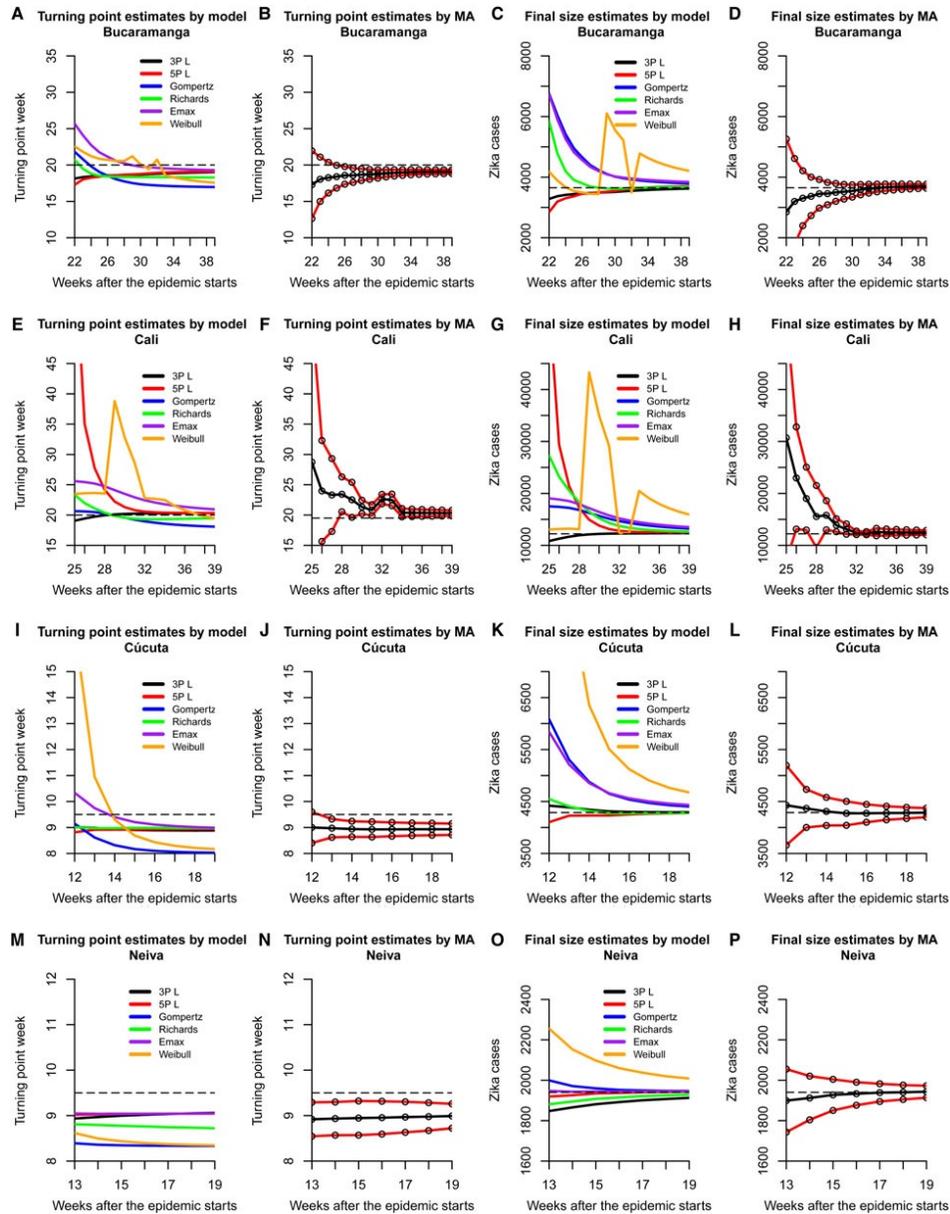


Figura 5.11: Estimación de parámetros para el acmé y tamaño final de la epidemia, a partir de los modelos no lineales bajo estudio (estimación puntual), a partir del promedio de modelo (estimación puntual e IC 95%) por ciudad. Las líneas discontinuas representan los valores observados. La escala de tiempo en todas las figuras presenta la última semana en el periodo de estimación.

Anexo 5

Salidas de los ejemplos de aplicación del Capítulo 3

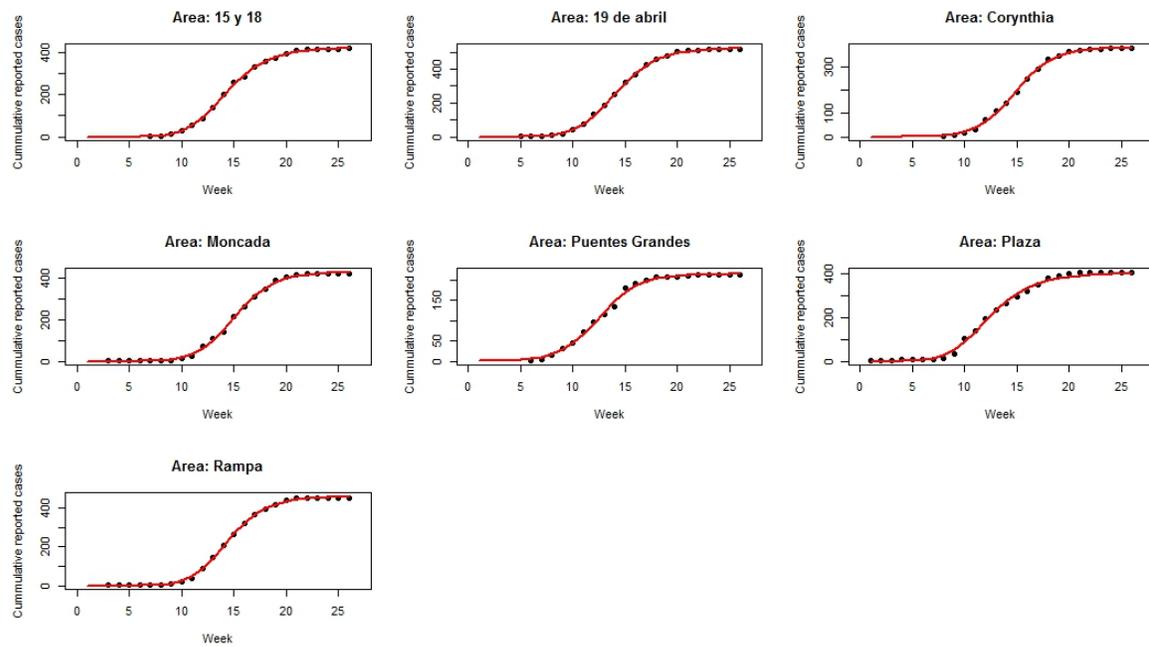


Figura 5.12: Número acumulado de casos observados y ajustados por el modelo de Richards asumiendo respuesta normal para las siete áreas de salud del municipio Plaza, La Habana, afectadas por el brote de dengue del 2006.

Tabla 5.6: Estimación de los parámetros y la matriz de correlación de los efectos aleatorios para los modelos con α, η, k (a) y α, η, γ (b) como efectos aleatorios.

a					b				
Parámetro	Estimación (s.e)				Parámetro	Estimación (s.e)			
Efectos fijos:					Efectos fijos:				
α	408.52 (33.68)				α	408.61 (33.72)			
k	0.4826 (0.054)				k	0.4803 (0.060)			
γ	0.9084 (0.079)				γ	0.9106 (0.090)			
η	13.5347 (0.3946)				η	13.5334 (0.3856)			
Varianza					Varianza				
σ^2	39.6327 (4.65)				σ^2	40.5256			
Efectos aleatorios		Matriz de correlación			Efectos aleatorios		Matriz de correlación		
$SD(b_{i1})$	87.78 (23.55)	α	k		$SD(b_{i1})$	87.9001 (23.55)	α	γ	
$SD(b_{i2})$	0.028 (0.012)	k	0.333		$SD(b_{i3})$	0.039 (0.012)	γ	0.364	
$SD(b_{i4})$	1.0162 (0.27)	η	0.557	0.957	$SD(b_{i4})$	0.9887 (0.27)	η	0.570	0.952

Tabla 5.7: Prueba de razón de verosimilitud para el modelo con matriz de varianza-covarianza Ψ definida positiva contra el modelo con matriz de varianza-covarianza bloque-diagonal.

Modelo	AIC	BIC	Log Ver.	Test	Razón V.	p -value
Modelo 1 (Ψ definida positiva)	1065.17	1097.99	-521.58			
Model 2 (Ψ Bloque-diagonal)	1064.28	1091.13	-523.14	1 vs 2	3.1132	0.2108

Anexo 6

Gráficos de diagnóstico del modelo no lineal con efectos mixtos

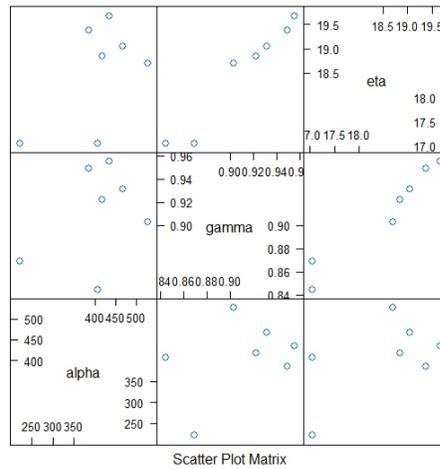


Figura 5.13: Gráfico de dispersión para las estimaciones de los efectos aleatorios del modelo con α , η y γ como efectos aleatorios.

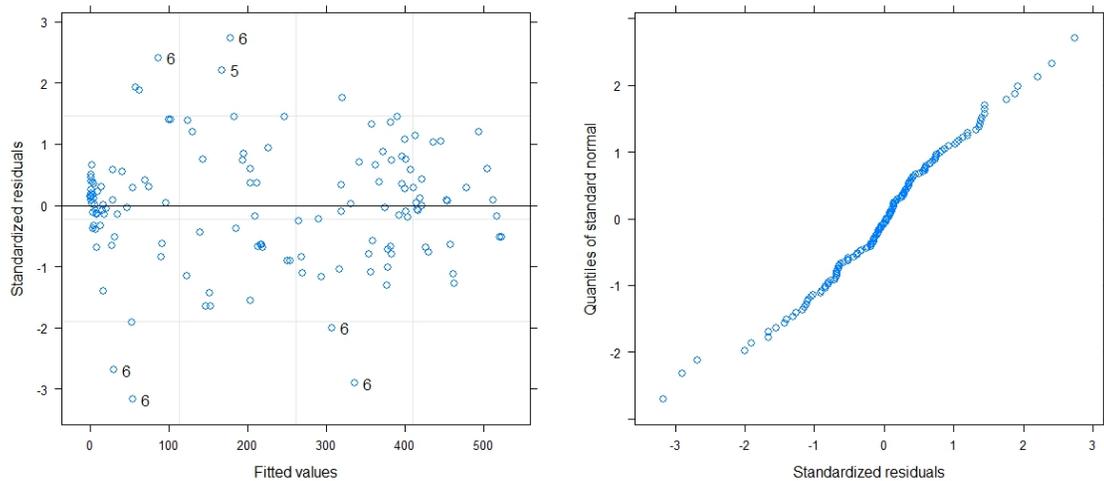


Figura 5.14: Gráfico de dispersión de los errores estandarizados (Panel izquierdo) y gráfico de la probabilidad normal de los residuos (Panel derecho) para el modelo con Ψ bloque diagonal.

Anexo 7

Códigos en SAS y R para la aplicación del Capítulo 3

A: Ajustando modelos para cada área específica

En SAS, estos modelos pueden ser estimado utilizando la rutina PROC NLMIXED

```
**** Health Area 15 y 18;
proc nlmixed data=d;
parms alpha=416, k=0.4856 , gamma=1 eta=13.8628 s2=10 ;
mu = alpha*((1+k*exp(-gamma*k*(semana-eta)))**(-1/k));
model acu15y18 ~ normal(mu,s2);
estimate 'R0' exp(gamma*(19/7));
run;

**** Health Area 19 abril;
proc nlmixed data=d;
parms alpha=519, k=1, gamma=1 eta=14 s2=1;
mu = alpha*((1+k*exp(-gamma*k*(semana-eta)))**(-1/k));
model acu19abril ~ normal(mu,s2);
estimate 'R0' exp(gamma*(19/7));
run;

**** Health Area Corynthia;
proc nlmixed data=d;
parms alpha=378, k=1, gamma=1 eta=14 s2=1;
mu = alpha*((1+k*exp(-gamma*k*(semana-eta)))**(-1/k));
model acucorinthy ~ normal(mu,s2);
estimate 'R0' exp(gamma*(19/7));
run;

**** Health Area Moncada;
proc nlmixed data=d;
parms alpha=425, k=1, gamma=1 eta=14 s2=1;
mu = alpha*((1+k*exp(-gamma*k*(semana-eta)))**(-1/k));
model acumoncada ~ normal(mu,s2);
```

```

estimate 'R0' exp(gamma*(19/7));
run;
**** Health Area Puentes Grandes;
proc nlmixed data=d;
parms alpha=215, k=1, gamma=1 eta=14 s2=1;
mu = alpha*((1+k*exp(-gamma*k*(semana-eta)))**(-1/k));
model acupgrande ~ normal(mu,s2);
estimate 'R0' exp(gamma*(19/7));
run;
**** Health Area Plaza;
proc nlmixed data=d;
parms alpha=399.14 k=0.49 gamma=0.86 eta=12 s2=10.4814 ;
mu = alpha*((1+k*exp(-gamma*k*(semana-eta)))**(-1/k));
model acuplaza ~ normal(mu,s2*s2);
estimate 'R0' exp(gamma*(19/7));
run;
**** Health Area Plaza;
proc nlmixed data=d;
parms alpha=455, k=1, gamma=1 eta=14 s2=10;
mu = alpha*((1+k*exp(-gamma*k*(semana-eta)))**(-1/k));
model acurampa ~ normal(mu,s2);
estimate 'R0' exp(gamma*(19/7));
run;

```

O adicionando la sentencia BY en la rutina PROC NLMIXED

```

proc nlmixed TECH=TRUREG data=d ;
by areas;
parms alpha=400, k=0.63, gamma=0.25 eta=13 s=1000;
mu = alpha*((1+k*exp(-gamma*k*(week-eta)))**(-1/k));
model cumcases ~ normal(mu,s);
estimate 'R0' exp(gamma*(19/7));
predict mu out=predicted2;
ODS output ParameterEstimates=Parameterestimates2;

```

```
run;
```

En R, estos modelos pueden ser estimado utilizando la función nlsList ()

```
####Data
areasplaza2006mixed<-read.delim2("areasplaza2006mixed.txt")
datmixed.new<-groupedData( cumcases ~ week | areas,data = areasplaza2006mixed)
# ####fit individual models
fit.Fix<-nlsList(cumcases~alpha/((1+k*exp(-gamma*k*(week-eta)))^(1/k)),data=datmixed.new,
start=list(alpha=406.5163,k=0.6897,gamma=0.6652, eta=13.8964)
summary(fit.Fix)
####Graph individual fitted models
plot(augPred(fit.Fix),layout=c(3,3),aspect=0.8,col=1, col.line=2,pch=20)
```

B: Abordaje con modelos con efectos mixtos

En SAS, se pueden ajustar estos modelos utilizando la rutina NLMIXED

```
*Mixed model with random effects: alpha, k, eta and Cov matrix positive-definite;
proc nlmixed TECH=TRUREG data=d;
parms alpha0=406.7452, k0=0.5176, gamma=0.8627 eta0=13.5590 g11=87 g22=0.03 g33=1 g21=0
g31=0 g32=1 s2=100;
alpha=alpha0+u1;
k=k0+u2;
eta=eta0+u3;
mu = alpha*((1+k*exp(-gamma*k*(week-eta)))*(-1/k));
model cumcases ~ normal(mu,s2);
random u1 u2 u3~normal([0,0,0],[g11*g11,g21,g22*g22,g31,g32,g33*g33]) subject=areas
out=efectosaleatorio;
estimate 'R0' exp(gamma*(19/7));
predict mu out=output_random1;
run;
```

En R, se requiere el uso de la función nlme()

```
library (nlme)
```

```

#### Model with no random effects
fit.mixed0<-nls(cumcases~alpha/((1+k*exp(-gamma*k*(week-eta)))^(1/k)),
data = datmixed.new,start=list(alpha=402.36,k=0.7704,gamma=0.8966,eta=13.8608))
summary(fit.mixed0)
#### Mixed model with 1 random effect: alpha
fit.mixed1<-nlme(cumcases~alpha/((1+k*exp(-gamma*k*(week-eta)))^(1/k)),
data = datmixed.new,
fixed=alpha+k+gamma+eta ~ 1,random=alpha~ 1,start=c(406.52,0.6897,0.6652,13.8964),
method="ML")
summary(fit.mixed1)
#### Mixed model with 2 random effects: alpha+k and Cov matrix positive-definite
fit.mixed21<-nlme(cumcases~alpha/((1+k*exp(-gamma*k*(week-eta)))^(1/k)),
data = datmixed.new, fixed=alpha+k+gamma+eta ~ 1,random=alpha+k ~ 1,
start=c(402.36,0.7704,0.8966, 13.8608),method="ML")
summary(fit.mixed21)
#### Mixed model with 2 random effects: alpha+gamma and Cov matrix positive-definite
fit.mixed22<-nlme(cumcases~alpha/((1+k*exp(-gamma*k*(week-eta)))^(1/k)),data = datmixed.new,
fixed=alpha+k+gamma+eta ~ 1,random=alpha+gamma ~ 1,start=c(402.36,0.7704,0.8966,13.8608),
method="ML")
summary(fit.mixed22)
#### Mixed model with 2 random effects: alpha+eta and Cov matrix positive-definite
fit.mixed23<-nlme(cumcases~alpha/((1+k*exp(-gamma*k*(week-eta)))^(1/k)),
data = datmixed.new, fixed=alpha+k+gamma+eta ~ 1,random=alpha+eta ~ 1,
start=c(402.36,0.7704,0.8966, 13.8608), method="ML")
summary(fit.mixed23)
#### Mixed model with 3 random effects: alpha+k+eta and Cov matrix positive-definite
fit.mixed31<-nlme(cumcases~alpha/((1+k*exp(-gamma*k*(week-eta)))^(1/k)),
data = datmixed.new, fixed=alpha+k+gamma+eta ~ 1,random=alpha+k+eta ~ 1,
start=c(402.36,0.7704,0.8966, 13.8608), method="ML")
summary(fit.mixed31)
#### Mixed model with 3 random effects: alpha+k+gamma and Cov matrix positive-definite
fit.mixed32<-nlme(cumcases~alpha/((1+k*exp(-gamma*k*(week-eta)))^(1/k)),
data = datmixed.new, fixed=alpha+k+gamma+eta ~ 1,random=alpha+k+gamma ~ 1,

```

```

start=c(402.36,0.7704,0.8966,13.8608), method="ML")
summary(fit.mixed32)
#### Mixed model with 3 random effects: alpha+eta+gamma and Cov matrix positive-definite
fit.mixed33<-nlme(cumcases~alpha/((1+k*exp(-gamma*k*(week-eta)))^(1/k)),
data = datmixed.new, fixed=alpha+k+gamma+eta ~ 1,random=alpha+gamma+eta ~ 1,
start=c(402.36,0.7704,0.8966, 13.8608),method="ML")
summary(fit.mixed33)
#### Mixed model with 4 random effects: alpha+k+eta+gamma and Cov matrix positive-definite
fit.mixed4<-nlme(cumcases~alpha/((1+k*exp(-gamma*k*(week-eta)))^(1/k)),
data = datmixed.new, fixed=alpha+k+gamma+eta ~ 1,random=alpha+k+gamma+eta ~ 1,
start=c(402.36,0.7704,0.8966, 13.8608),method="ML")
summary(fit.mixed4)

```

Códigos en R para algunas tablas y figuras de la aplicación

```

#### Pairs plot for the random-effects estimate, Figure 5.12
pairs(fit.mixed31)
#### Model31 with Covariance matrix Block diagonal
fit.mixed31update <- update( fit.mixed31, random = pdBlocked(list(k + eta ~ 1, alpha ~ 1)) )
summary(fit.mixed31update)
#### Likelihood ratio test, Table 3.10
anova(fit.mixed31,fit.mixed31update)
#### Scatter plot of standardized residual versus fitted values, panel izquierdo Figura 5.13
plot(fit.mixed31update,id=0.05,adj=-1)
#### Normal plot of standardized residual, panel derecho Figura 5.13
qqnorm(fit.mixed31update)
#### plot of the augmented predictions, Figure 3.2
plot(augPred(fit.mixed31update,level = 0:1), key=list(text = list(c("Population",
"Individual (Area)","Observed")), lines = list(type=c("l","l","p"), pch=20, col=c(2,4,1),
lty=c(1,2,1)),columns=3), layout=c(3,3), aspect=0.8, col=1, col.line=c(2,4),lty=c(1,2),
pch=20, xlab="Weeks", ylab="Cumulative number of reported cases")

```

Anexo 8

Archivo DESCRIPTION del paquete *DengueRT*

Package: DengueRT
Type: Package
Title: Parameter Estimates and Real-Time Prediction of a Single Dengue Outbreak
Version: 1.0.1
Date: 2016-04-19
Author: Carlos Sebrango, Lizet Sanchez, Ziv Shkedy, Ewoud De Troyer
Maintainer: Carlos Sebrango <sebrango@uniss.edu.cu>
Description: Provides functions for parameter estimation and real-time predictions of a single dengue outbreak taking into account model uncertainty using model averaging.
LazyData: true
Depends: R (>= 2.10), ggplot2, nlme, drc
License: GPL-3

Archivo NAMESPACE del paquete *DengueRT*

```
export(allmodels,allmodelpredict,changetimeFSTP)
import(nlme,ggplot2,drc)
S3method(summary,dengue)
S3method(plot,dengue)
importFrom("grDevices", "dev.new")
importFrom("graphics", "legend", "lines", "plot")
importFrom("stats", "AIC", "D", "coef", "nls", "predict", "qt")
```

Anexo 9

Vista del sitio Web, descargas y ayuda del paquete *DengueRT*

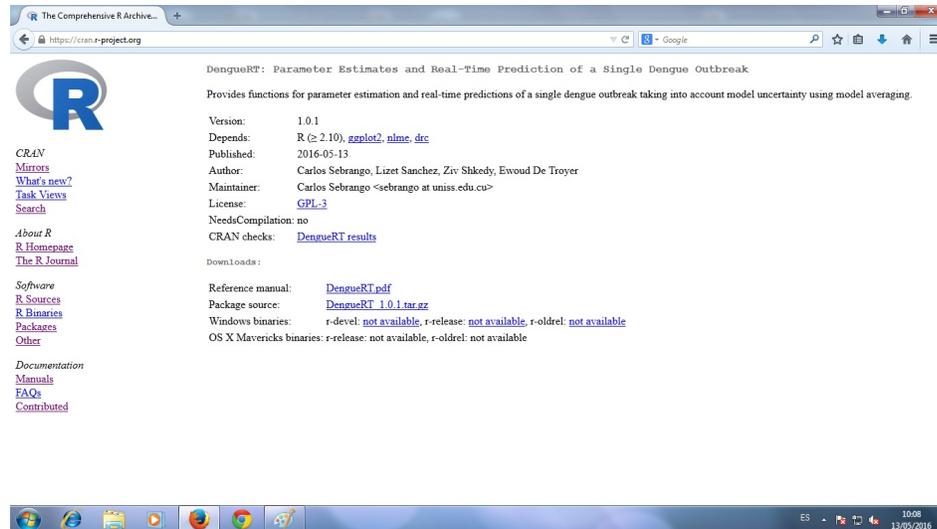


Figura 5.15: Vista de la página Web del paquete *DengueRT* en el CRAN.

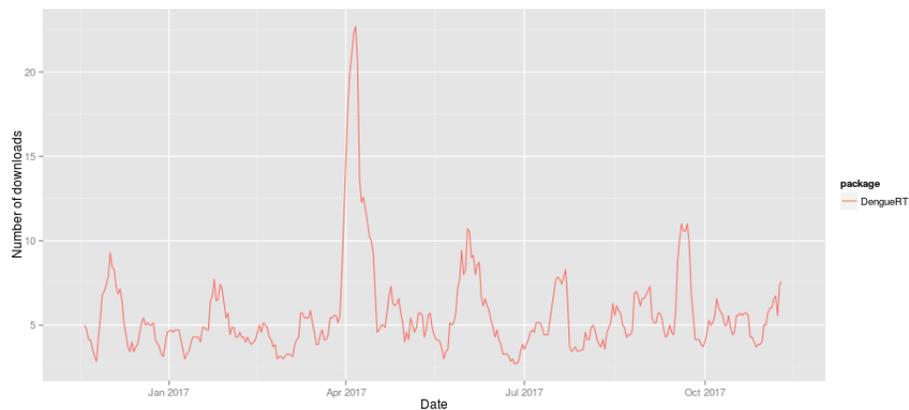


Figura 5.16: Descargas del paquete *DengueRT* en el año 2017.

En las siguientes páginas se muestra la ayuda del paquete *DengueRT* tal y como aparece en el sitio Web <https://cran.r-project.org/web/packages/DengueRT/>.

Package ‘DengueRT’

May 13, 2016

Type Package

Title Parameter Estimates and Real-Time Prediction of a Single Dengue Outbreak

Version 1.0.1

Date 2016-04-19

Author Carlos Sebrango, Lizet Sanchez, Ziv Shkedy, Ewoud De Troyer

Maintainer Carlos Sebrango <sebrango@uniss.edu.cu>

Description Provides functions for parameter estimation and real-time predictions of a single dengue outbreak taking into account model uncertainty using model averaging.

LazyData true

Depends R (>= 2.10), ggplot2, nlme, drc

License GPL-3

Repository CRAN

Repository/R-Forge/Project dengueparmet

Repository/R-Forge/Revision 10

Repository/R-Forge/DateTimeStamp 2016-05-12 13:58:47

Date/Publication 2016-05-13 00:50:06

NeedsCompilation no

R topics documented:

allmodelpredict	2
allmodels	4
changetimeFSTP	8
dengueoutbreak1	9
dengueoutbreak2	10
DengueRT	10
plot	11
summary	13

Index	16
--------------	-----------

allmodelpredict	<i>Real-time predictions of the final size and the turning point at the end of the epidemic, the incidence and the cumulative number of cases in future observations.</i>
-----------------	---

Description

This function provides real-time predictions of the final size and the turning point at the end of the epidemic for each built-in model and model averaged, as well as the incidence and the cumulative number of cases in future observations. Also this function, when all the built-in models are used, gives the AIC of each model, the model averaged weights and predicted incidence and cumulative cases.

Usage

```
## Object of the S3 class dengue
allmodelpredict(inc,time,pred,start=NULL,model)
```

Arguments

inc,time	Vector of equal length specifying incidence (number of reported cases per time unit) and time interval (from the start of outbreak).
pred	Number of observation in which the incidence and the cumulative number of cases will be predicted.
start	A list with the starting values of the model to be used for fitting the data. If model="all" the input must be a list of a list with the starting values of Richards, 3P logistic, Sigmoid Emax, Gompertz, Weibull and 5P logistic model parameters. By default, the initial values are provided by self-starting functions (see argument start in allmodels function).
model	The nonlinear model to be used for fitting the data. Built-in models are "Richards", "Logistic3P", "SigmEmax", "Gompertz", "Weibull" and "Logistic5P" (see argument model in allmodels function). If model = "all" the parameter estimate will be done taking into account all built-in models via model averaging.

Value

An object with the parameter estimate for each built-in model and model averaged estimate for final size and turning point of outbreak. It is a list:

Incidence	All the available incidences
Time	All the available time points
PredTime	Period of time for which the prediction is required
AIC	The AIC for each built-in model and model averaged
tTable	A table with parameter estimates and t test. It is not available when all the model are used.

Weights	Model averaged weights. It is not available when is used only one built-in model
FinalSize	95% confidence interval and point estimate of the final size of outbreak for each built-in model and model averaged estimate
TurningPoint	95% confidence interval and point estimate of the turning point of outbreak for each built-in model and model averaged estimate
Predict	Predicted cumulative cases for each built-in model
PredictMA	Predicted cumulative cases for model averaged
PredInc	Predicted incidence for each built-in model
PredMAInc	Predicted incidence for model averaged
function.type	Name of the function
model.type	models used to estimate

Generic functions such as plot and summary have methods to show the results of the fit

Author(s)

Carlos Sebrango, Lizet Sanchez, Ziv Shkedy

References

- K. Burnham, D. R. Anderson, Model Selection and Multimodel Inference: A Practical Information - theoretic Approach, 2nd Edition, Springer-Verlag, New York, 2002.
- G. Claeskens, N. L. Hjort, Model selection and model averaging, Cambridge University Press, 2008.

Examples

```
## Not run:
## (data example 1)
data("dengueoutbreak1")

## Using only the information until time point 20
## Prediction of the final size and turning point at the end of epidemic,
## the incidence and the cumulative number of cases in the observation
## number 36 using only the Richards model
allmodelpredict(dengueoutbreak1$Incidence[1:20],dengueoutbreak1$Time[1:20],36,
model = "Richards")

## Using only the information until time point 22
## Now using all built-in model, Prediction of the final size and turning point
## at the end of epidemic, the incidence and the cumulative number of cases
## in the observation number 30
allmodelpredict(dengueoutbreak1$Incidence[1:22],dengueoutbreak1$Time[1:22],30,
model = "all")
## End(Not run)

## (data example 2)
data("dengueoutbreak2")
## Using only the information until time point 18, Prediction of the final size and
```

```

## turning point at the end of epidemic, the incidence and the cumulative number
## of cases in the observation number 31 using only the 3P logistic model

allmodelpredict(dengueoutbreak2$Incidence[1:18],dengueoutbreak2$Time[1:18],31,
model = "logistic3P")

## Not run:
## Using only the information until time point 20
## Now using all built-in model, Prediction of the final size and turning point
## at the end of epidemic, the incidence and the cumulative number of cases
## in the observation number 30
allmodelpredict(dengueoutbreak2$Incidence[1:20],dengueoutbreak2$Time[1:20],34,
model = "all")
## End(Not run)

```

allmodels	<i>Parameter estimate for each built-in model and model averaged estimate for final size and turning point of outbreak.</i>
-----------	---

Description

This function gives the parameter estimate for each built-in model and model averaged estimate for final size and turning point of outbreak. Also this function, when all the built-in models are used, gives the AIC of each model, the model averaged weights and predicted incidence and cumulative cases. This function is used retrospectively, that is, when all the data are available.

Usage

```

## Object of the S3 class dengue
allmodels(inc,time,start=NULL,model)

```

Arguments

inc,time	Vector of equal length specifying incidence (number of reported cases per time unit) and time interval (from the start of outbreak).
start	<p>A list with the starting values of the model to be used for fitting the data. If model="all" the input must be a list of a list with the starting values of Richards, 3P logistic, Sigmoid Emax, Gompertz, Weibull and 5P logistic model parameters. By default, the initial values are provided by self-starting functions.</p> <ul style="list-style-type: none"> • 3P Logistic model uses SSlogis selfstart function from the stats package. • Sigmoid Emax model uses LL.4 selfstart function from the drc package. • Gompertz model uses SSgompertz selfstart function from the stats package. • Weibull model uses SSweibull selfstart function from stats package. • 5P Logistic model uses LL.5 and LL.4 selfstart functions from drc package. • Richards model uses SSlogis selfstart function and some ideas of SSRichards selfstart function from NRAIA package now archived.

model

The nonlinear model to be used for fitting the data. Built-in models are "Richards", "Logistic3P", "SigmEmax", "Gompertz", "Weibull" and "Logistic5P". If model = "all" the parameter estimate will be done taking into account all built-in models via model averaging. For all built-in model, $C(t)$ represents the cumulative number of reported cases at time t and also the turning point (η) and the final size of epidemic (α) are parameters in the models. The model expressions of each built-in model are given as follow:

- The Richards model is given by the expression:

$$C(t) = \alpha / [1 + k \cdot \exp(-k \cdot \gamma \cdot (t - \eta))]^{1/k}$$

Where

γ : Per capita growth rate or intrinsic growth rate.

k : Exponent of the deviation from standard logistic curve.

- The three parameter logistic model (3P Logistic), proposed by Verhulst, is given by the expression:

$$C(t) = \alpha / [1 + \exp(-\gamma \cdot (t - \eta))]$$

Where

γ : Per capita growth rate or intrinsic growth rate.

- The Sigmoid Emax model is given by the expression:

$$C(t) = \beta + (t^n)(\alpha - \beta) / (t^n + \eta^n)$$

Where

n : Slope factor or Hill factor.

β : Lower asymptote.

- The Gompertz model is given by the expression :

$$C(t) = \beta + (\alpha - \beta) \cdot \exp(-\exp(-\gamma(t - \eta)))$$

Where

γ : Per capita growth rate or intrinsic growth rate.

β : Lower asymptote.

- The Weibull model is given by the expression:

$$C(t) = \alpha + (\beta - \alpha) \cdot \exp(-(t/\eta)^k)$$

Where

k : Shape factor.

β : Lower asymptote.

- The five parameter logistic model (5P Logistic) is given by the expression:

$$C(t) = \alpha + (\beta - \alpha) / [1 + (2^{1/g} - 1) \cdot (t/\eta)^k]^g$$

Where

g: Asymmetry factor.
 k: Shape factor.
 beta: Lower asymptote.

Value

An object with the parameter estimate for each built-in model and model averaged estimate for final size and turning point of outbreak. It is a list:

Incidence	All the available incidences
Time	All the available time points
AIC	The AIC for each built-in model and model averaged
tTable	A table with parameter estimates and t test. It is not available when all the model are used.
Weights	Model averaged weights. It is not available when is used only one built-in model
FinalSize	95% confidence interval and point estimate of the final size of outbreak for each built-in model and model averaged estimate
TurningPoint	95% confidence interval and point estimate of the turning point of outbreak for each built-in model and model averaged estimate
Predict	Predicted cumulative cases for each built-in model
PredictMA	Predicted cumulative cases for model averaged
PredInc	Predicted incidence for each built-in model
PredMAInc	Predicted incidence for model averaged
function.type	Name of the function
model.type	models used to estimate

Generic functions such as plot and summary have methods to show the results of the fit

Author(s)

Carlos Sebrango, Lizet Sanchez, Ziv Shkedy

References

- K. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*, 2nd Edition, Springer-Verlag, New York, 2002.
- J. MacDougall, Analysis of dose responses Studies: Emax model, in: N. Ting (Ed.), *Dose Finding in Drug Development*, Statistics for Biology and Health, Springer New York, pp. 127, 2006.
- G. Claeskens, N. L. Hjort, *Model selection and model averaging*, Cambridge University Press, 2008.
- D. Ratkowsky, *Handbook of nonlinear regression models*, Marcel Dekker, New York, 1990.
- F. Richards, A flexible growth function for empirical use, *Journal of Experimental Botany* 10 (29), pp 290-301, 1959.
- Y.H. Hsieh, Temporal trend and regional variability of 2001-2002 multiwave DENV-3 epidemic in Havana City: did Hurricane Michelle contribute to its severity?, *Tropical Medicine and International Health*, Vol. 18, no. 7, pp 830-838, 2013.

A. Tsoularis, J. Wallace, Analysis of logistic growth models, *Mathematical Biosciences*, Vol. 179, no. 1, pp 21-55, 2002.

J. Liao, R. Liu, Re-parameterization of five-parameter logistic function, *Journal of Chemometrics* 23 (5), pp 248-253, 2009.

Examples

```
## (data example 1)
data("dengueoutbreak1")
## Not run:
## Parameter estimate for Richards model (for Incidence data example 1)
allmodels(dengueoutbreak1$Incidence,dengueoutbreak1$Time,
start=list(alpha=375,k=2.38,gamma=0.76,eta=16), model="Richards")
## End(Not run)
## or
p1<-allmodels(dengueoutbreak1$Incidence,dengueoutbreak1$Time, model="Richards")
## summary function for a allmodels object
summary(p1)
## plot function for a allmodels object
plot(p1,which=c(1,2))

## Not run:
## Parameter estimate for each built-in model and model averaged
##estimate for final size and turning point of outbreak

allmodels(dengueoutbreak1$Incidence,dengueoutbreak1$Time,
start=list(list(alpha=375,k=2.38,gamma=0.76,eta=16),
list(alpha=375,gamma=1,eta=16),list(alpha=380,eta=13,beta=7,n=5),
list(alpha=380,eta=20,beta=0,gamma=1),list(alpha=410,eta=12,beta=11,k=3),
list(alpha=380,beta=4,g=1,eta=13,k=15)), model="all")
## or
allmodels(dengueoutbreak1$Incidence,dengueoutbreak1$Time,model="all")

## (data example 2)
data("dengueoutbreak2")
# Parameter estimate for 3P Logistic model
allmodels(dengueoutbreak2$Incidence,dengueoutbreak2$Time,
start=list(alpha=375,gamma=1,eta=16), model="logistic3P")
## or
allmodels(dengueoutbreak2$Incidence,dengueoutbreak2$Time,model="logistic3P")

## Parameter estimate for each built-in model and model averaged estimate
##for final size and turning point of outbreak
##for Incidence data example 2
allmodels(dengueoutbreak2$Incidence,dengueoutbreak2$Time,
start=list(list(alpha=355,k=1,gamma=1,eta=14),
list(alpha=355,gamma=1,eta=14), list(alpha=355,eta=13,beta=10,n=6),
list(alpha=355,eta=11,beta=20,gamma=1),list(alpha=355,eta=12,beta=22,k=3),
list(alpha=355,beta=15,g=1,eta=13,k=10)),model="all")
## or
allmodels(dengueoutbreak2$Incidence,dengueoutbreak2$Time,model="all")
## End(Not run)
```

changetimeFSTP	<i>Changes over time of the final size and turning point estimates for each built-in model and model averaged.</i>
----------------	--

Description

This function provides the changes over time of the final size and turning point estimates (for specific models and model average) from the time point required until the last time point available. The output of the function are two tables, one for final size and the other for turning point. the rows represent the data used for the estimation of the model parameters.

Usage

```
## Object of the S3 class dengue
changetimeFSTP(inc, time, ini, start = NULL)
```

Arguments

<code>inc, time</code>	Vector of equal length specifying incidence (number of reported cases per time unit) and time interval (from the start of outbreak).
<code>ini</code>	Time point from which is required to see the changes over time of the final size and turning point estimates
<code>start</code>	A list with the starting values of the model to be used for fitting the data. If <code>model="all"</code> the input must be a list of a list with the starting values of Richards, 3P logistic, Sigmoid Emax, Gompertz, Weibull and 5P logistic model parameters. By default, the initial values are provided by self-starting functions.

Value

An object with the changes over time of the final size and turning point estimates for specific models and model average. It is a list:

<code>Incidence</code>	All the available incidences
<code>Time</code>	All the available time points
<code>Period</code>	Time period required to see the changes over time of the final size and turning point estimates
<code>FSchangetime</code>	A Table with the changes over time of the final size for the time period required
<code>TPchangetime</code>	A Table with the changes over time of the turning point for the time period required
<code>function.type</code>	Name of the function
<code>model.type</code>	models used to estimate

Generic functions such as `plot` and `summary` have methods to show the results of the fit.

Author(s)

Carlos Sebrango, Lizet Sanchez, Ziv Shkedy

References

K. Burnham, D. R. Anderson, Model Selection and Multimodel Inference: A Practical Information-theoretic Approach, 2nd Edition, Springer-Verlag, New York, 2002.
G. Claeskens, N. L. Hjort, Model selection and model averaging, Cambridge University Press, 2008.

Examples

```
## Not run:
## (data example 2)
data("dengueoutbreak2")

##Time point from which is required to see the changes over time
##of the final size and turning point estimates:20

t2<-changetimeFSTP(dengueoutbreak2$Incidence,dengueoutbreak2$Time,20)

##summary function for changetimeFSTP object
summary(t2)

## plot function for changetimeFSTP object
plot(t2,which=c(5,6))

## End(Not run)
```

dengueoutbreak1	<i>Incidence data of the 2001/2002 single dengue outbreak in the health area "26 de julio", Havana City, Cuba.</i>
-----------------	--

Description

This data set gives the incidence (Reported dengue cases) in each time point (in weeks) of the 2001/2002 single dengue outbreak in the health area "26 de julio" of the Playa municipality, Havana City, as compiled by the Institute of Tropical Medicine "Pedro Kouri", Havana City, Cuba.

Usage

```
data("dengueoutbreak1")
```

Format

A data frame with 36 observations on the following 2 variables.

Incidence a numeric vector, representing the incidence of the outbreak.

Time a numeric vector, representing the time vector in weeks.

Source

Institute of Tropical Medicine "Pedro Kouri", Havana City, Cuba.

dengueoutbreak2	<i>Incidence data of the 2001/2002 single dengue outbreak in the health area "J. R. Ramirez", Havana City, Cuba.</i>
-----------------	--

Description

This data set gives the incidence (Reported dengue cases) in each time point (in weeks) of the 2001/2002 single dengue outbreak in the health area "J. R. Ramirez" of the Playa municipality, Havana City, as compiled by the Institute of Tropical Medicine "Pedro Kouri", Havana City, Cuba.

Usage

```
data("dengueoutbreak2")
```

Format

A data frame with 34 observations on the following 2 variables.

Incidence a numeric vector, representing the incidence of the outbreak.

Time a numeric vector, representing the time vector in weeks.

Source

Institute of Tropical Medicine "Pedro Kouri", Havana City, Cuba.

DengueRT	<i>Parameter Estimates and Real-Time Prediction of a Single Dengue Outbreak</i>
----------	---

Description

The DengueRT package uses the incidence data from a single dengue outbreak and provides functions to estimate the final size, the turning point of the epidemic and to conduct a real-time prediction for these parameters using several nonlinear models taking into account model uncertainty using model averaging. Graphical tools for a visualization of the results are also included.

Details

Package: DengueRT
 Type: Package
 Version: 1.0.1
 Date: 2016-04-19
 License: GPL-3

Author(s)

Carlos Sebrango, Lizet Sanchez, Ziv Shkedy, Ewoud De Troyer

Maintainer: Carlos Sebrango <sebrango@uniss.edu.cu>

References

- K. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*, 2nd Edition, Springer-Verlag, New York, 2002.
- J. MacDougall, Analysis of dose responses Studies: Emax model, in: N. Ting (Ed.), *Dose Finding in Drug Development*, Statistics for Biology and Health, Springer New York, pp. 127, 2006.
- G. Claeskens, N. L. Hjort, *Model selection and model averaging*, Cambridge University Press, 2008.
- D. Ratkowsky, *Handbook of nonlinear regression models*, Marcel Dekker, New York, 1990.
- F. Richards, A flexible growth function for empirical use, *Journal of Experimental Botany* 10 (29), pp 290-301, 1959.
- Y.H. Hsieh, Temporal trend and regional variability of 2001-2002 multiwave DENV-3 epidemic in Havana City: did Hurricane Michelle contribute to its severity?, *Tropical Medicine and International Health*, Vol. 18, no. 7, pp 830-838, 2013.
- A. Tsoularis, J. Wallace, Analysis of logistic growth models, *Mathematical Biosciences*, Vol. 179, no. 1, pp 21-55, 2002.
- J. Liao, R. Liu, Re-parameterization of five-parameter logistic function, *Journal of Chemometrics* 23 (5), pp 248-253, 2009.

plot

Plots for a visualization of the results

Description

plot is a generic function used to produce plots for a visualization of the results of the functions: allmodels, allmodelpredict and changetimeFSTP from the S3 class dengue. The function invokes particular methods of the S3 class dengue.

Usage

```
## S3 method for class 'dengue'  
plot(x, which = c(1, 2), xlab = "", ...)
```

Arguments

- x An object for which a plot is desired. It could be any of the following objects: allmodels, allmodelpredict and changetimeFSTP
- which Different graph options, from 1 to 6.

- Option 1 gives a plot with the cumulative number of cases and fitted models for each built-in model and model averaged. This option is only available for `allmodels` and `allmodelpredict` objects.
- Option 2 gives a plot with the incidence and fitted models for each built-in model and model averaged. This option is only available for `allmodels` and `allmodelpredict` objects.
- Option 3 gives a plot with the final size 95% confidence interval and point estimate for each built-in model and model averaged. This option is only available for `allmodels` and `allmodelpredict` objects when all models are used.
- Option 4 gives a plot with the turning point 95% confidence interval and point estimate for each built-in model and model averaged. This option is only available for `allmodels` and `allmodelpredict` objects when all models are used.
- Option 5 gives a plot with the change over time of the final size estimates for each built-in model and model averaged. This option is only available for `changetimeFSTP` objects.
- Option 6 gives a plot with the change over time of the turning point estimates for each built-in model and model averaged. This option is only available for `changetimeFSTP` objects.

`xlab` A title for the x axis. It is not necessary for options 3 y 4.

`...` Additional arguments affecting the plot produced. Arguments to be passed to methods, such as graphical parameters (lines, segment, etc)

Details

When the which options 1, 2, 5 and 6 are used, the plot displays all fitted models with the following colours: (Richards (red), 3P logistic (green), Sigmoid Emax (dark blue), Gompertz (clear blue), Weibull (pink) and 5P logistic (yellow))and the fitted model via model averaging (gray colour). The which options 1, 2, 3 and 4 are only used with the `allmodels` and `allmodelpredict` objects. The which options 5 and 6 are only used with the `changetimeFSTP` object.

Author(s)

Carlos Sebrango, Lizet Sanchez, Ziv Shkedy

References

K. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*, 2nd Edition, Springer-Verlag, New York, 2002.

G. Claeskens, N. L. Hjort, *Model selection and model averaging*, Cambridge University Press, 2008.

Examples

```
## Not run:
## (data example 1)
data("dengueoutbreak1")
```

```

## plot function for a allmodels object
##Richards model
ee<-allmodels(dengueoutbreak1$Incidence,dengueoutbreak1$Time, model="Richards")
plot(ee,which=c(1,2))

## (data example 2)
data("dengueoutbreak2")
##using all built-in models
ee2<-allmodels(dengueoutbreak2$Incidence,dengueoutbreak2$Time, model="all")
plot(ee2,which=c(1,2,3,4))

##plot function for allmodelpredict object
##3P logistic model (Incidence data example 2)
## information available to the time point 20
## observation to predict: 30
ee3<-allmodelpredict(dengueoutbreak2$Incidence[1:20],dengueoutbreak2$Time[1:20],30,
model="logistic3P")
plot(ee3,which=c(1,2))

##using all built-in models (Incidence data example 2)
## information available to the time point 22
## observation to predict: 30
ee4<-allmodelpredict(dengueoutbreak2$Incidence[1:22],dengueoutbreak2$Time[1:22],30,
model="all")
plot(ee4,which=c(1,2,3,4))
## End(Not run)

##plot function for changetimeFSTP object
## See example changetimeFSTP function

```

summary

Result summaries of the functions

Description

summary is a generic function used to produce result summaries of the functions: allmodels, allmodelpredict and changetimeFSTP from the S3 class dengue. The function invokes particular methods of the S3 class dengue.

Usage

```

## S3 method for class 'dengue'
summary(object, ...)

```

Arguments

object	An object for which a summary is desired. It could be any of the following objects: allmodels, allmodelpredict and changetimeFSTP.
...	Additional arguments affecting the summary produced.

Value

The form of the value returned by summary depends on the object of the S3 class dengue. In the case of allmodels returns the AIC, model averaged Weights and model averaged estimates for the final size and turning point of outbreak. If the object is allmodelpredict to the returned value is added a prediction of the incidence and cumulative cases in the required future observation. When the object is changetimeFSTP function the returned values are two tables with the changes over time of the final size and turning point for the time period required.

Author(s)

Carlos Sebrango, Lizet Sanchez, Ziv Shkedy, Ewoud DE Troyer

References

K. Burnham, D. R. Anderson, Model Selection and Multimodel Inference: A Practical Information-theoretic Approach, 2nd Edition, Springer-Verlag, New York, 2002.
G. Claeskens, N. L. Hjort, Model selection and model averaging, Cambridge University Press, 2008.

Examples

```
## Not run:
## (data example 1)
data("dengueoutbreak1")

## summary function for a allmodels object
##Richards model (Incidence data example 1)
ee<-allmodels(dengueoutbreak1$Incidence,dengueoutbreak1$Time, model="Richards")
summary(ee)

## (data example 2)
data("dengueoutbreak2")
##using all built-in models (Incidence data example 2)
ee2<-allmodels(dengueoutbreak2$Incidence,dengueoutbreak2$Time, model="all")
summary(ee2)

##summary function for allmodelpredict object
##3P logistic model (Incidence data example 2)
## information available to the time point 20
## observation to predict: 30
ee3<-allmodelpredict(dengueoutbreak2$Incidence[1:20],dengueoutbreak2$Time[1:20],30,
model="logistic3P")
summary(ee3)

##using all built-in models (Incidence data example 1)
## information available to the time point 22
## observation to predict: 30
ee4<-allmodelpredict(dengueoutbreak1$Incidence[1:22],dengueoutbreak1$Time[1:22],30,
model="all")
summary(ee4)
## End(Not run)
```

```
##summary function for changetimeFSTP object  
## See example changetimeFSTP function
```

Index

*Topic **datasets**

dengueoutbreak1, [9](#)

dengueoutbreak2, [10](#)

allmodelpredict, [2](#)

allmodels, [4](#)

changetimeFSTP, [8](#)

dengueoutbreak1, [9](#)

dengueoutbreak2, [10](#)

DengueRT, [10](#)

plot, [11](#)

summary, [13](#)