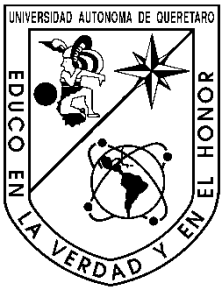

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

FACULTAD DE INGENIERÍA

MAESTRÍA EN CIENCIAS EN INGENIERÍA MATEMÁTICA

COMPARACIÓN DE MÉTODOS DE INFERENCIA ESTADÍSTICA SOBRE PARÁMETROS DE SISTEMAS DINÁMICOS NO LINEALES DE PRIMER ORDEN

TESIS

QUE COMO PARTE DE LOS REQUISITOS PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS EN INGENIERÍA MATEMÁTICA

PRESENTA:

L.M.A. BERNARDO CHÁVEZ CASTILLO

DIRIGIDA POR:

Dr. EDUARDO CASTAÑO TOSTADO

CENTRO UNIVERSITARIO, QUERÉTARO, QRO.
NOVIEMBRE DE 2014

Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría en Ciencias en Ingeniería Matemática

COMPARACIÓN DE MÉTODOS DE INFERENCIA ESTADÍSTICA SOBRE PARÁMETROS DE SISTEMAS DINÁMICOS NO LINEALES DE PRIMER ORDEN

TESIS

Que como parte de los requisitos para obtener el Grado de
Maestro en Ciencias en Ingeniería Matemática

Presenta:
L.M.A. Bernardo Chávez Castillo

Dirigido por:
Dr. Eduardo Castaño Tostado

Dr. Eduardo Castaño Tostado — Presidente — Firma

M. en C. Enrique Crespo Baltar — Secretario — Firma

Dr. Víctor Manuel Armando Aguirre Torres — Vocal — Firma

M. en C. Sara Silva Hernández — Suplente — Firma

Dr. Russell James Bowater — Suplente — Firma

Dr. Aurelio Domínguez González
Director de la Facultad

Dr. Irineo Torres Pacheco
Director de Investigación y Posgrado

Centro Universitario
Querétaro, Qro.
Noviembre de 2014

Resumen

Gran parte de los fenómenos naturales poseen una alta complejidad por lo que resulta difícil proveer de modelos que los reproduzcan fielmente. El uso de ecuaciones diferenciales ordinarias no lineales ha logrado proveer de suficiente flexibilidad para reproducir dinámicas complejas debido a su manejo de interacciones entre variables de estado e inclusión de parámetros específicos del fenómeno. En este sentido, es necesario estimar los parámetros del modelo considerando aquellas fuentes de incertidumbre que impiden una total captura del fenómeno. Son pocos los métodos de estimación de parámetros que permiten capturar incertidumbre, de entre ellos se destacan los métodos estadísticos Bayesianos, específicamente, aquellos que utilizan cómputo intensivo mediante muestreo de tipo Markov Chain Monte Carlo (MCMC). Este trabajo estudia numéricamente la estimación de parámetros a través de técnicas Bayesianas, contrastando los algoritmos Metropolis-Hastings (MH) y population-based MCMC (PB), en su eficiencia, mediante simulación de datos con ruido y el manejo de modelos de diversa complejidad de la familia presa-depredador (Lotka-Volterra). En sistemas cíclicos estacionarios el método PB mostró mejores características que las mostradas por el MH, en la estimación tanto de parámetros como en la solución del sistema. Se estudiaron los efectos en la estimación Bayesiana debidos al ruido observacional y a dinámicas con un componente no aleatorio pero caótico, particularmente el llamado sistema de Lorenz. Se observó que la estimación PB Bayesiana, aplicada a un sistema caótico, arroja un estimador que recupera la tendencia general del sistema independientemente del nivel del ruido muestral y de la complejidad del modelo (caótico o no caótico). Finalmente se presenta un par de aplicaciones que engloban el proceso de modelado y estimación de parámetros ante problemas de identificabilidad del modelo. Se concluye que los métodos de estimación no pueden lidiar directamente con este problema pero arrojan pistas en la estimación de parámetros (multimodalidad y/o intervalos de probabilidad muy amplios) sobre la presencia del mismo; pistas que no son tomadas en cuenta generalmente en la práctica aplicada del modelaje, dejan claro que no siempre los modelos teóricos se desempeñan adecuadamente, por lo que se debe tener precaución al utilizarlos y eventualmente optar por modelos empíricos.

(**Palabras clave:** MCMC, Sistema de Lorenz, Caos)

Summary

Natural phenomena possess a high complexity which makes difficult to provide models that reproduce them faithfully. The use of non linear ordinary differential equations has permitted to provide of sufficient flexibility to reproduce complex dynamics due to its managing of interactions between state variables and the incorporation of specific parameters of the phenomenon. In this sense, it is necessary to estimate these parameters considering those uncertainty sources that prevent a total apprehension of the phenomenon. There are few methods of parameter estimation that allow the capture of uncertainty. In this work are outlined the statistical Bayesian methods, specifically, those that use intensive computing by means of Markov Chain Monte-Carlo (MCMC) sampling. This work studies numerically the parameter estimation across two Bayesian tactics, contrasting the algorithms Metropolis-Hastings (MH) and population-based MCMC (PB), in their efficiency, by means of simulating data with different noise levels and using diverse complexity models of the so called prey-predator family (Lotka-Volterra).In cyclical stationary systems the PB method showed better characteristics than those showed by the MH, in the parameter estimation as in the system solution. The effects due to the observational noise and dynamics with a non random but chaotic component were studied in the Bayesian estimation, particularly the so called Lorenz system. It was observed that the Bayesian estimation PB, applied to a chaotic system, throws an estimator that recovers the general trend of the system independently of the sampling noise level and of the complexity of the model (chaotic or no). Finally, we present a couple of applications that cover the process of modeling and the parameter estimation facing identifiability problems of the model. It is concluded that these estimation methods cannot directly confront identification problems but only give clues in the parameter estimation (multimodality and / or wide probability intervals) about its presence; if these clues are taken into account in applied modeling they clarify the extent of applicability of theoretical models.

(**Key words:** MCMC, Lorenz System, Chaos)

Dedicatorias

A mi familia por su desinteresado apoyo.

A mi Mary Birds por estar conmigo en las buenas y las malas.

A todos aquellos que ya no están y a los que han llegado a mi vida...

Agradecimientos

Agradezco a todas las personas e instituciones que directa o indirectamente ayudaron a la realización de éste trabajo de investigación.

A mis padres por su apoyo y desinteresado amor a lo largo de mi vida.

A mis compañeros y amigos por apoyarme con sus conocimientos y su sincera amistad.

A los miembros del jurado que a pesar de sus tantas ocupaciones, sus comentarios permitieron enriquecer este trabajo.

A mi asesor, el Dr. Castaño por su paciencia y apoyo en la dirección de esta tesis, y como profesor, al abrirme las puertas a nuevos conocimientos.

Finalmente, y no menos importante, agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por haberme otorgado la beca para poder llevar a cabo el presente trabajo hasta su término.

A todos, gracias.

Índice

Índice de figuras

Índice de cuadros

1. Introducción

La búsqueda de modelos matemáticos explicativos de los fenómenos presentes en el campo de la ciencia, tiene como principal dificultad la de plasmar el fenómeno en un modelo que logre reproducir el comportamiento cualitativo esperado del mismo. Sin embargo, es evidente que muchos de esos fenómenos de interés presentan una alta complejidad lo que resulta de una gama amplia de fuentes de incertidumbre que impide la captura total del fenómeno en un modelo. Es así que se han enfocado los esfuerzos por proponer técnicas que permitan estimar los parámetros de modelos altamente complejos, sin dejar de lado la captura de incertidumbre, para así, proporcionar una manera de reproducir y predecir fenómenos (Schaber y Klipp, 2011).

Una primera respuesta para tratar fenómenos con dinámicas tortuosas son los modelos dinámicos no lineales, ya que poseen la capacidad de reproducir tales dinámicas a partir del manejo de interacciones entre las variables de estado del fenómeno y inclusión de parámetros específicos. Por lo anterior, es evidente que el uso de ecuaciones diferenciales no lineales provee de un marco teórico y práctico lo bastante flexible y natural para tratar varios tipos de problemas.

El proponer un modelo dinámico que reproduzca cierto fenómeno, implica entre varias cosas, definir un sistema de ecuaciones diferenciales dependiente de las variables de estado observadas $\mathbf{x}$, y de un vector de parámetros θ, que se supondrá rige la dinámica observada. Tales ecuaciones y parámetros reflejan nuestras hipótesis a partir de la información disponible del fenómeno. Suponiendo que las ecuaciones están bien especificadas, el problema se reduce finalmente a la estimación de los parámetros. En este sentido, se puede afirmar que la estimación de parámetros surge como un intento de ajustar un modelo a un conjunto de datos experimentales, $\mathbf{y}$, provenientes de la práctica y que se encuentran sujetos a un cierto error experimental, e. Esta estimación de parámetros se ha abordado generalmente como un problema de optimización, específicamente, como la de minimizar una distancia dada entre el modelo y los datos experimentales; sin embargo, este enfoque se ve limitado muchas de las veces por el cómputo intensivo para el cálculo de la solución y de una definición adecuada de la distancia para la estimación de los parámetros.

El problema de estimación de parámetros ha dado paso a un mundo diverso de metodologías de estimación. Sin embargo, no todas ellas resuelven las dificultades presentes, por lo que muchas son técnicas muy específicas. La minimización de la suma de errores al cuadrado (comúnmente denominada "*mínimos cuadrados*") es un método de estimación de parámetros tradicional, sin embargo, este método resulta ser altamente ineficiente ante valores atípicos, lo que se acrecienta al enfrentarse a problemas no lineales (Ramsay *et al*., 2007). El proceso de estimación parte de una muestra de datos observados en el campo, sujetos a perturbaciones del medio y del fenómeno mismo, es por ello que se requiera de un método de estimación

que, además, permita cuantificar tal incertidumbre. Métodos tradicionales impiden obtener tal cuantificación, sin embargo, aproximaciones más estadísticas como la estimación por máxima verosimilitud y métodos Bayesianos, si bien presentan problemas de estimación ante sistemas complejos no lineales como cualquier método (Varziri *et al*., 2008), permiten cuantificar mínimamente la incertidumbre del fenómeno, situación que requiere trabajo adicional en el caso de los mínimos cuadrados no lineales.

El presente trabajo se enfoca en la estimación de parámetros a partir de técnicas Bayesianas en sistemas de ecuaciones diferenciales ordinarias no lineales, debido a su habilidad en la captura de incertidumbre muestral. En una primera parte se exponen algunos de los enfoques en lo que a estimación de parámetros se refiere. Desde métodos puramente numéricos hasta el uso de métodos altamente teóricos, los enfoques presentan ventajas y desventajas que exponen su potencial ante los problemas que la práctica presenta. Dentro de esta gran variedad de métodos de estimación, debemos destacar la importancia de reconocer la incertidumbre del contexto; pues ello se verá reflejado en nuestros resultados y conclusiones. Como se mencionó, nuestro estudio se enfoca en los métodos Bayesianos, específicamente, métodos de estimación que utilizan muestreo de tipo *Markov Chain Monte Carlo* (*MCMC*), que no son más que técnicas de simulación de procesos estocásticos que poseen distribuciones de probabilidad conocidas y pueden ser usadas para aproximar la distribución final de los parámetros del modelo.

Los métodos MCMC realizan inferencia estadística generando funciones de probabilidad de los parámetros a ser estimados. Mediante muestreo repetido, dado el modelo $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \boldsymbol{\theta})$, la inferencia Bayesiana busca la distribución de densidad posterior $p(\boldsymbol{\theta}|\mathbf{y})$ de los parámetros, partiendo de una distribución a priori $p(\boldsymbol{\theta})$; la cual posee toda aquella información disponible del parámetro $\boldsymbol{\theta}$ por parte del usuario. Los algoritmos *Metropolis-Hastings* y *population-based MCMC*, como algoritmos representativos de los métodos de muestreo, serán estudiados en su habilidad de realizar inferencia estadística sobre parámetros de modelos dinámicos a partir de datos experimentales. Mediante la simulación de datos con ruido y el manejo de un mismo fenómeno con modelos de diversa complejidad, se analizará la eficiencia de los métodos MCMC ante la estimación de parámetros. Si bien es cierto que se puede distinguir en un fenómeno, ya sea una dinámica simple o compleja, no es claro su efecto en la estimación de parámetros, es por ello que a partir de un par de modelos de tipo *presa-depredador* (*Lotka-Volterra*), modelo biológico que describe la competencia natural entre dos especies por los recursos disponibles, se podrán caracterizar aquellas propiedades que pongan en evidencia las limitaciones y fortalezas de los algoritmos MCMC, justificando finalmente, el uso de los mismos a lo largo de la investigación.

Posteriormente y tras haber contrastado los métodos Bayesianos, resultando el algoritmo *population-based MCMC* como el más adecuado para la estimación, se prosigue con la

identificación de los efectos del ruido y de la complejidad de la dinámica sobre la estimación. Para ello, se utilizarán datos con distintos niveles de ruido experimental y dinámicas con componentes caóticos, es decir, un componente no aleatorio pero altamente complejo. Se presentan dos vertientes, un sistema con un estado no caótico pero con dinámica oscilatoria estable y finalmente un sistema en un estado caótico, cada uno con distintos niveles de ruido muestral. Lo anterior reflejará el efecto de un modelo bien especificado pero caótico y sujeto a ruido muestral, sobre las distribuciones posteriores de los parámetros, y por tanto en las soluciones estimadas.

El trabajo se concluye presentando un par de aplicaciones que engloban el proceso de modelado y estimación de parámetros en un sistema, así como el problema de identificabilidad del mismo. La primera aplicación corresponde a un proceso de modificación de un modelo propuesto que explica los datos provenientes de un biorreactor, donde se ve reflejada la carencia de ajuste en las distribuciones posteriores de los parámetros, y donde la modificación a partir de ello nos lleva hasta un modelo que recupera la dinámica experimental de la mejor manera posible. Finalmente, a partir de datos correspondientes a un proceso de hidrólisis, se contrastará un par de modelos, reflejando así los peligros de aferrarse a modelos basados en supuestos insostenibles, que nos lleve a resultados incorrectos, por lo que se debería optar por otro tipo de modelos menos teóricos.

El resultado del proceso de modelaje y estimación, es un modelo parametrizado que permite realizar predicciones cuantitativas sobre el comportamiento del fenómeno de manera rápida y a bajo costo (Schaber y Klipp, 2011), maximizando la obtención de información útil.

2. Antecedentes

Introducción

El desarrollo científico implica la búsqueda de modelos matemáticos que permita reproducir y manipular los fenómenos de acuerdo a los intereses científicos que se tengan. Chou y Void (2009) resumen lo que consideran el proceso de modelado en sistemas dinámicos y lo dividen en nueve partes que no sólo contemplan el proponer un modelo y la estimación de sus parámetros, sino que es un proceso que incluye la selección de datos experimentales, una recopilación de información sobre el fenómeno, la especificación de supuestos y simplificaciones, posteriormente la estimación de parámetros así como también el diagnóstico, validación y refinamiento del modelo para su posterior aplicación. Es así que el modelaje de cualquier fenómeno requiere de la realización de hipótesis específicas y el uso de conceptos rigurosos del área. En este mismo sentido, la formulación del modelo matemático requiere considerar las limitaciones físicas del fenómeno así como visualizar todo el entramado presente en el fenómeno. Es durante este proceso de construcción que se hace evidente, muchas de las veces, nuestra falta de conocimiento y habilidades para plasmar el fenómeno en un modelo que logre mostrar el comportamiento cualitativo esperado del fenómeno (Schaber y Klipp, 2011).

El modelaje plantea, como objetivo, capturar esencialmente la complejidad del fenómeno y también permitir un cómputo eficiente, esto último debido a que no necesariamente se pueden obtener soluciones analíticas de muchos sistemas lo que hace necesario el uso del cálculo computacional para su aproximación. La estimación de parámetros nace como un intento de ajustar un modelo a un conjunto de datos experimentales, es cierto que el modelo no es único, y también es cierto que existen diversas metodologías para tal fin que logran optimizar la estimación.

Notación

Un modelo dinámico se representa por medio de un sistema de ecuaciones diferenciales, para nuestros fines se están considerando únicamente ecuaciones diferenciales ordinarias de primer orden no lineales, con una estructura general dada por la expresión (2.1), donde $\mathbf{x}(t)$ representa el vector de estados del sistema (variables de salida) en el tiempo t, $\dot{\mathbf{x}}$ el vector de derivadas de los estados del sistema, $\boldsymbol{\theta}$ el vector de parámetros desconocidos a estimar y $\mathbf{x}_0$ el vector de condiciones iniciales.

$$\begin{cases} \dot{\mathbf{x}}(t) = F(t, \mathbf{x}(t), \boldsymbol{\theta}), \quad t \in [0, T] \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \tag{2.1}$$

Generalmente, sólo se tiene acceso a datos observados sujetos a un cierto error experimental, en este sentido, dadas m variables de salida **x**, medidas en tiempos t_{ij}, $i = 1, .., N_j$ y $j \in \{1, ..., m\}$, donde N_j representa el número de observaciones disponibles por variable, entonces, las mediciones experimentales al tiempo t_{ij} se definen como y_{ij}, sujetas a un error experimental e_{ij}, es decir, $y_{ij} = x(t_{ij}) + e_{ij}$.

2.1. Estimación de parámetros

El proceso de estimación de parámetros se ha manejado comúnmente como un problema de optimización, en el cual se requiere de minimizar cierta función objetivo que define una cierta distancia entre los datos experimentales y el modelo. Cuando se emplea la distancia Euclidiana entonces nos enfrentamos a un problema de mínimos cuadrados tradicional. Es entonces que el uso de integración numérica intensiva es requerida para dar solución a este problema de optimización, sin embargo, este proceso suele tener un alto costo computacional; y es así que una gama de métodos de estimación han sido desarrollados como alternativas, de donde se destacan dos grupos importantes, los métodos basados en gradiente y los algoritmos estocásticos de búsqueda (Chou y Voit, 2009) . Dentro de los primeros existen algoritmos muy conocidos como el Gauss-Newton y el Levenberg–Marquardt. Dada la complejidad de muchos sistemas, estos algoritmos corren el riesgo de quedar atrapados en óptimos locales lo que hace a su eficiencia altamente dependiente del grado de complejidad y los valores iniciales dados a los parámetros. Por otro lado, los algoritmos estocásticos de búsqueda, que incluyen; cómputo evolutivo, recocido simulado, métodos adaptativos estocásticos, etc; poseen un alto potencial de encontrar óptimos globales o al menos regiones cercanas a los mismos.

Tradicionalmente, métodos como los mínimos cuadrados (2.2) han sido utilizados para obtener una estimación de los parámetros de algún sistema a partir de observaciones dadas. Generalizaciones como los mínimos cuadrados no lineales (MCNL) han sido utilizadas para trabajar en modelos más complejos bajo un funcionamiento similar. A partir de algoritmos numéricos como el Runge-Kutta se aproxima la solución del sistema dado un conjunto de parámetros y una condición inicial, posteriormente se introducen los valores ajustados del paso anterior en un algoritmo de optimización que actualiza nuevamente los parámetros a estimar.

$$\widehat{\theta} = \underset{\theta}{arg\,min} \sum_{i=1}^{n} (y_i - x(t_i))^2 \tag{2.2}$$

Tal procedimiento presenta diversos problemas, por un lado, requiere de cómputo intensivo puesto que se utiliza integración numérica en cada actualización de parámetros y de condición inicial (Ramsay *et al*., 2007), por otro, al enfrentarse a sistemas no lineales o de problemas de especificación, hace que se deba explorar de manera intensiva el espacio de

parámetros, lidiando con mínimos locales y por tanto, con un alto costo en tiempo de ejecución (Brunel, 2008). Ante este hecho, la exactitud de las aproximaciones numéricas puede ser un problema. El número de parámetros puede incrementarse al considerar condiciones iniciales adicionales necesarias para resolver el sistema lo que se traduce, en muchos casos, en falta de información proveniente de los datos. Finalmente, la estimación por MCNL, es puntual, por lo que no se tiene acceso a estimación de intervalos, lo que se traduce en un trabajo extra de cómputo para su obtención (Ramsay *et al*., 2007).

2.2. Enfoques frecuentista y Bayesiano

Desde el punto de vista estadístico, los enfoques frecuentista y Bayesiano se destacan en el sentido de proveer herramientas para la captura y explicación de incertidumbre en los parámetros. Suponiendo que las mediciones realizadas están sujetas a un cierto error experimental de tipo aditivo **e**, descrito por alguna distribución de probabilidad, por ejemplo Normal $\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, podemos representar tales datos en función de los estados del sistema $\mathbf{x}(\mathbf{t}, \boldsymbol{\theta})$, y el ruido experimental por medio de la expresión siguiente:

$$\mathbf{y} = \mathbf{x}(\mathbf{t}, \boldsymbol{\theta}) + \mathbf{e}$$

En este sentido, considerando la función de densidad que rige la probabilidad de observar los datos experimentales **y**, dados ciertos valores de los parámetros $\boldsymbol{\theta}$, que gobiernan el sistema, tenemos entonces la función de densidad de probabilidad p (de los datos) denominada función de verosimilitud que tiene la forma de la expresión (2.3).

$$p(\mathbf{y}|\boldsymbol{\theta}) = \prod_{j=1}^{m} \prod_{i=1}^{N_j} p(\mathbf{y}_j(\mathbf{t}_i)|\boldsymbol{\theta}) \tag{2.3}$$

Los enfoques Bayesiano y frecuentista difieren a partir de este punto en algunos aspectos, por un lado el enfoque frecuentista se centra en la búsqueda de aquellos valores de los parámetros que logren describir los datos experimentales bajo un cierto grado de aceptabilidad, es decir, se busca un estimador para el cual dado un cierto valor umbral ν, se cumpla que $p(\mathbf{y}|\boldsymbol{\theta}) > \nu$. En este sentido se utilizan ciertos niveles de significancia y del calculo de valores críticos.

El enfoque Bayesiano tradicional, a partir de un conjunto de observaciones experimentales $\mathcal{D} = \{\mathbf{y}, \mathbf{t}\}$, tiene como objetivo, obtener un modelo que relacione de algún modo **t** con **y**. Definamos el conjunto de todos los modelos posibles por $\mathcal{M} = \{\mathcal{M}_1, ..., \mathcal{M}_k\}$, donde cada modelo $\mathcal{M}_k$ está definido por un conjunto de parámetros $\boldsymbol{\theta}_k$.

Cada modelo tiene asociado un conjunto de parámetros por lo que la verosimilitud del modelo $\mathcal{M}_k$ con parámetros $\boldsymbol{\theta}_k \in \Theta$ (Θ espacio de todos los posibles valores de parámetros $\boldsymbol{\theta}_k$) que da origen a las observaciones **y**, es la densidad de probabilidad $p\left(\mathcal{D}|\boldsymbol{\theta}_k, \mathcal{M}_k\right)$ de manera similar a la ecuación (2.3). En este mismo sentido se asigna una cierta densidad de probabilidad a priori a los parámetros correspondientes al modelo, es decir $p\left(\boldsymbol{\theta}_k|\mathcal{M}_k\right)$, de donde se obtiene una verosimilitud integrada, que corresponde al valor esperado de la verosimilitud de los datos respecto a la a priori del parámetro (Girolami, 2008).

$$p\left(\mathcal{D}|\mathcal{M}_k\right) = \int_{\boldsymbol{\theta}_k \in \Theta} p\left(\mathcal{D}|\boldsymbol{\theta}_k, \mathcal{M}_k\right) p\left(\boldsymbol{\theta}_k|\mathcal{M}_k\right) d\boldsymbol{\theta}_k$$

De esta forma y del teorema de Bayes (2.5), obtenemos a partir de una distribución a priori, y la verosimilitud de los datos, la distribución a posteriori (2.4) de los parámetros $\boldsymbol{\theta}_k$ asociados al modelo $\mathcal{M}_k$.

$$p\left(\boldsymbol{\theta}_k|\mathcal{D}, \mathcal{M}_k\right) = \frac{p\left(\mathcal{D}|\boldsymbol{\theta}_k, \mathcal{M}_k\right) p\left(\boldsymbol{\theta}_k|\mathcal{M}_k\right)}{p\left(\mathcal{D}|\mathcal{M}_k\right)} \tag{2.4}$$

$$p\left(\boldsymbol{\theta}|\mathbf{y}\right) = \frac{p\left(\mathbf{y}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)}{\int_{\boldsymbol{\theta} \in \Theta} p\left(\mathbf{y}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta}} \tag{2.5}$$

Mientras el enfoque Bayesiano tiene como objetivo muestrear de la distribución posterior de los parámetros $p\left(\boldsymbol{\theta}|\mathbf{y}\right)$ y determinar un intervalo probabilidad usando la densidad de probabilidad a posteriori, el enfoque frecuentista se basa en maximizar la función de verosimilitud de los datos (2.3). Un punto compartido con el enfoque Bayesiano es que también se puede adicionar información a priori del fenómeno a los parámetros, ello se denomina estimador de Máxima A Posteriori, *MAP*, (Vanlier *et al*., 2013).

La estimación por máxima verosimilitud (MV), si bien es un método muy popular, gracias a sus características asintóticas, presenta problemas ante sistemas no lineales complejos (Varziri *et al*., 2008). El método Bayesiano presenta problemas similares. Mediante el uso del teorema de Bayes (2.5), para obtener la distribución a posteriori de los parámetros se requiere de la introducción de información a priori. Si se tiene un espacio multidimensional de los parámetros, la a posteriori no es manejable por integración numérica, por lo que se requiere del uso de métodos de tipo MCMC, ya mencionados con anterioridad. Sin embargo, tales métodos requieren nuevamente de integración numérica iterativa. La a posteriori obtenida de esta forma no posee una forma cerrada, y dada su complejidad, se pueden presentar problemas de convergencia por parte de los métodos MCMC. El método Bayesiano y el de mínimos cuadrados requieren de buenas condiciones iniciales para los parámetros. Si se carece de tales condiciones, entonces surgen problemas de convergencia. Métodos de disparo múltiple (multiple

shooting methods) y métodos de barrera o punto interior (barrier or interior point), han mostrado ser más estables que métodos basados en gradiente, además de tener más oportunidad de converger, aún con condiciones iniciales pobres de los parámetros. Sin embargo, aún cuando requieren de relativamente pocas iteraciones para converger, requieren de mucho poder de cómputo (Gugcomoushvil y Klaassen, 2012).

Moles *et al*. (2003) aborda problemas de programación no lineal en la estimación de parámetros en reacciones bioquímicas, que involucran restricciones no lineales, afirma que métodos de optimización locales tradicionales (basados en gradiente) fallan en llegar a una solución satisfactoria en estos problemas, además de explicar que ciertos tipos de algoritmos estadísticos, particularmente estrategias evolutivas, resuelven tales problemas de manera satisfactoria. Si bien los métodos determinísticos proveen de cierto nivel de seguridad de que se pueda converger al óptimo global, no es así en general para ciertos problemas, además de que el costo computacional crece rápidamente al aumentar el tamaño del problema (variables, parámetros, restricciones, etc.). Por otro lado, los métodos estocásticos, al involucrar un elemento aleatorio, poseen una garantía teórica débil de converger a la solución global, sin embargo pueden localizar una vecindad de la solución global con una eficiencia relativamente alta, pero con el costo de que tal optimalidad pudiera no ser garantizada. Cual sea el caso, todos estos métodos de estimación permiten agregar información, por parte del usuario, sobre la topología de los espacios de los parámetros, al proponer distribuciones de probabilidad a priori, en el caso probabilístico, o valores iniciales para el caso determinista.

2.3. Alternativas

Alternativas a los mínimos cuadrados, MV y métodos de muestreo Bayesianos, son los métodos de dos pasos (two-step methods), los cuales siguen un enfoque diferente: en un primer paso se calcula la solución del sistema a partir de métodos no paramétricos por medio de una función auxiliar, finalmente, en un segundo paso se estiman los parámetros a partir de la minimización de alguna distancia dada (Brunel 2008). Históricamente, se han usado métodos de suavizamiento para obtener una función auxiliar que emule los datos experimentales, se destacan aproximaciones como splines mínimos cuadrados (Varah, 1982), splines cúbicos (Madar *et al*., 2003), regresión polinomial local y polinomios de Lagrange, por ejemplo.

Este método de suavizamiento ha sido considerado por Ramsay y Silverman (1997) en su esquema de análisis de datos funcionales ADF, (FDA, functional data analysis), que transforma los datos en funciones mediante splines cúbicos suavizados, así como en un método posterior denominado suavizamiento generalizado (Ramsay *et al*., 2007), donde el suavizamiento de la solución y la estimación de parámetros es considerada conjuntamente. El uso de estimadores no paramétricos es motivado por la simplicidad computacional de los métodos de

estimación, adicionalmente desde el punto de vista funcional, permite usar información a priori sobre las soluciones, como positividad o acotamiento donde es difícil explorar estrictamente la forma paramétrica del sistema (Brunel, 2008) .

El primero de los métodos correspondientes a la familia de métodos de dos pasos, aquí mencionados, corresponde al descrito en Ramsay *et al*. (2007), denominado Suavizamiento Generalizado (SG). Considere un sistema de ecuaciones diferenciales ordinarias de la forma

$$\begin{cases} \dot{\mathbf{x}}(t) = F(\mathbf{x}(t), \boldsymbol{\theta}), \quad t \in [0, T] \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \tag{2.6}$$

Cuyas observaciones o datos experimentales está dadas por los estados del sistema junto con un error aditivo. El método de SG corresponde a la familia de métodos de colocación, por lo que expresa aproximaciones $\widehat{\mathbf{y}}_i$ de las observaciones $\mathbf{y}_i$ en términos de funciones base, en la forma

$$\widehat{\mathbf{y}}_j(t) = \sum_{k}^{K_j} c_{jk}\phi_{jk}(t) = \mathbf{c}_j^{'}\boldsymbol{\phi}_j(t)$$

donde K_i es el número de funciones base en el vector ϕ_i para asegurar flexibilidad para capturar la variación en la aproximación de la función $\mathbf{y}_j$ y sus derivadas. Dado que los parámetros $\mathbf{c}_j$ no son de principal interés, pero se requiere de su estimación, se denominan parámetros incómodos. Los parámetros incómodos son estimados al ser definidos como funciones implícitas $\widehat{\mathbf{c}}_{\mathrm{j}}(\boldsymbol{\theta}, \boldsymbol{\sigma}; \boldsymbol{\lambda})$ de parámetros estructurales en el sentido de que cada vez que $\boldsymbol{\theta}$ y $\boldsymbol{\sigma}$ son cambiados, un criterio de ajuste $J(\widehat{\mathbf{c}}|\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$ es re-optimizado con respecto a $\widehat{\mathbf{c}}$. La función de estimación $\widehat{\mathbf{c}}_{\mathrm{j}}(\boldsymbol{\theta}, \boldsymbol{\sigma}; \boldsymbol{\lambda})$ es regularizada al incorporar un término de penalización en J que controla el tamaño con que $\widehat{\mathbf{y}} = \widehat{\mathbf{c}}'\phi$ falla en satisfacer la ecuación diferencial. Tal cantidad de regularización es controlada por un parámetro de suavizamiento $\boldsymbol{\lambda}$. Un criterio de ajuste a los datos $H(\boldsymbol{\theta}, \boldsymbol{\sigma}|\boldsymbol{\lambda})$ es entonces optimizado con respecto a los parámetros estructurales.

Si definimos los errores asociados a la variable j, por $\mathbf{e}_j$, sea $g(\mathbf{e}_j|\boldsymbol{\sigma}_j)$, la densidad conjunta de esos errores condicionales al vector de parámetros $\boldsymbol{\sigma}_j$, entonces, definamos el criterio de ajuste a los datos $H(\boldsymbol{\theta}, \boldsymbol{\sigma}|\boldsymbol{\lambda})$, por medio de la log-verosimilitud,

$$H(\boldsymbol{\theta}, \boldsymbol{\sigma}|\boldsymbol{\lambda}) = -\sum_{j} \ln\{g(\mathbf{e}_j|\boldsymbol{\sigma}_j, \boldsymbol{\theta}, \boldsymbol{\lambda})\}$$

De forma similar, si se cuenta con suficientes datos, se puede definir un criterio compuesto de suma de cuadrados

$$H(\boldsymbol{\theta}|\boldsymbol{\lambda}) = \sum_{j} w_j \left\|\mathbf{y}_j - \widehat{\mathbf{x}}_j\right\|^2$$

La dependencia de H sobre $(\boldsymbol{\theta}, \boldsymbol{\sigma})$ es en dos sentidos, directa e implícitamente a través de $\widehat{\mathbf{c}}_{\mathrm{j}}(\boldsymbol{\theta}, \boldsymbol{\sigma}; \boldsymbol{\lambda})$ al definir el ajuste de $\widehat{\mathbf{y}}_i$. El parámetro $\boldsymbol{\lambda}$ de suavizamiento se ajusta utilizando algunas heurísticas numéricas y visuales, sin embargo, tal parámetro también puede ser estimado automáticamente por medio de una función $F(\boldsymbol{\lambda})$. La manera en cómo se sintoniza el parámetro $\boldsymbol{\lambda}$, es similar al uso de la información a priori en la estimación Bayesiana, sin embargo, tal estrategia Bayesiana puede llevar a expresiones no analíticas lo que se traduce en el uso de cómputo intensivo como el uso de técnicas de tipo MCMC. El método aquí expuesto lleva a expresiones analíticas para una eficiente optimización y también para una estimación de intervalos, que no es posible con métodos como mínimos cuadrados (Ramsay *et al*., 2007).

El sistema en (2.6), se puede expresar para cada variable j, como un operador de la forma

$$L_{j,\boldsymbol{\theta}} = \dot{\mathbf{x}}_j(t) - F(\mathbf{x}_j(t), \boldsymbol{\theta})$$

La magnitud con la cual un función $\widehat{\mathbf{x}}_j$ satisface el sistema $\dot{\mathbf{x}}(t) = F(\mathbf{x}(t), \boldsymbol{\theta})$ puede ser evaluada por una función de fidelidad

$$PEN_j = \int L_{j,\boldsymbol{\theta}} \left\{\widehat{\mathbf{x}}_j(t)\right\}^2 dt$$

De manera conjunta se define una medida compuesta de fidelidad por

$$PEN\left(\widehat{\mathbf{x}}|\mathbf{L}_{\boldsymbol{\theta}}, \boldsymbol{\lambda}\right) = \sum_j \lambda_j PEN_j\left(\widehat{\mathbf{x}}\right)$$

Finalmente, la función de ajuste a los datos y la ecuación de fidelidad se combinan para dar paso el criterio de ajuste interior

$$J(\mathbf{c}|\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\lambda}) = -\sum_j \ln\left\{g\left(\mathbf{e}_j|\boldsymbol{\sigma}_j, \boldsymbol{\theta}, \boldsymbol{\lambda}\right)\right\} + PEN\left(\widehat{\mathbf{x}}|\boldsymbol{\lambda}\right)$$

o de manera similar

$$J(\mathbf{c}|\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\lambda}) = \sum_j w_j \left\|\mathbf{y}_j - \widehat{\mathbf{x}}_j\right\|^2 + PEN_j\left(\widehat{\mathbf{x}}|\boldsymbol{\theta}, \boldsymbol{\lambda}\right)$$

de donde se obtiene la estimación de $\widehat{\mathbf{c}}$.

La minimización de una distancia entre estimadores no paramétricos y una familia paramétrica, es un mecanismo clásico de estimación no paramétrica (Estimador de Mínima Distancia, EMD). La diferencia entre el estimador de dos pasos y el estimador EMD es que el primero no minimiza directamente la distancia entre la función de regresión y un modelo paramétrico,

más bien la distancia entre las derivadas de la función de regresión y el modelo paramétrico (Brunel, 2008).

El método de dos pasos es más cercano al de la estimación EMD que el de suavizamiento generalizado anteriormente expuesto. El enfoque de suavizamiento generalizado encuentra una curva g que resuelva aproximadamente una ecuación diferencial, la cual es entonces cercana a la aproximación por splines de la solución de la EDO calculada por colocación:

$$\underset{g,\theta}{min} \sum_{i=1}^{n} |y_i - g(t_i)|_2^2 \text{ sujeto a } \|\dot{g} - F(\cdot, g, \theta)\|_2^2 < \epsilon$$

El problema de optimización es resuelto de forma iterativa, lo cual implica que el estimador no paramétrico es calculado de manera adaptativa con respecto al modelo paramétrico y a los datos, contrario al método de estimación a dos pasos, donde usa los datos una vez sin explotar el modelo paramétrico (Brunel, 2008).

En este sentido se menciona el uso de un estimador no paramétrico para el sistema (2.6) definido en el intervalo $[0, 1]$, sin pérdida de generalidad, que evite explotar el modelo paramétrico como lo hace el método anterior. Para ello, sea $\mathbf{x}^*$ la solución al sistema F correspondiente al vector de parámetros real del sistema $\boldsymbol{\theta}^*$. Definamos además, la norma

$$\|f\|_{q,w} = \left(\int_0^1 |f(t)|^q w(t)\, dt \right)^{1/q}$$

Donde $w(t)$ es una función positiva y definida en $[0, 1]$. El principio de la estimación en dos pasos es motivado por el hecho de que es bastante fácil construir estimadores consistentes de la solución $\mathbf{x}^*$ y de su derivada $\dot{\mathbf{x}}^*$. Mediante el uso de splines, funciones kernel o polinomios, se construye un estimador $\widehat{\mathbf{x}}_n$ de $\mathbf{x}^*$, del cual también se deriva un estimador para $\dot{\mathbf{x}}^*$. El criterio usado se define como

$$R_{n,w}^q(\boldsymbol{\theta}) = \left\| \dot{\widehat{\mathbf{x}}}_n - F\left(t, \widehat{\mathbf{x}}_n, \boldsymbol{\theta}\right) \right\|_{q,w}$$

Que minimiza la distancia entre el modelo paramétrico y su estimador, de donde se deriva el estimador (Brunel, 2008):

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_n &= arg\underset{\boldsymbol{\theta}}{min} R_{n,w}^q(\boldsymbol{\theta}) \\ \widehat{\boldsymbol{\theta}}_n &= arg\underset{\boldsymbol{\theta}}{min} \left(\int_0^1 \left| \dot{\widehat{\mathbf{x}}}_n - F\left(t, \widehat{\mathbf{x}}_n, \boldsymbol{\theta}\right) \right|^q w(t)\, dt \right)^{1/q} \end{aligned}$$

En este camino de la búsqueda de estimadores no paramétricos consistentes, Gugush-

vili y Klaassen (2012), a diferencia de Brunel (2008), que usa regresión por splines, plantean el uso de un estimador de x_j por funciones kernel,

$$\widehat{x}_j = \sum_{i=1}^{n} (t_i - t_{i-1}) \frac{1}{b} K\left(\frac{t - t_{i-1}}{b}\right) y_{ij}$$

Donde $K(\cdot)$ es una función kernel y b denota una ventana dependiente del tamaño muestral. Aquí, b representa un parámetro de regularización para ajustar la varianza y sesgo. Bajo este esquema, se define el estimador θ_n de θ,

$$\widehat{\theta}_n = \underset{\eta}{arg\,min} \int_0^1 \left\| \dot{\widehat{x}}(t) - F\left(\widehat{x}(t), \eta\right) \right\|^2 w(t)\, dt$$

Los métodos basados en dos pasos no se consideran como competidores de la estimación por mínimos cuadrados o el método Bayesiano. Puesto que éstos últimos métodos requieren de una buena condición inicial para la estimación, el método de dos pasos puede ser un buen complemento para dar una estimación razonable los parámetros iniciales. Una limitante del método es que requiere que todas las variables de estado estén disponibles (variables observadas), sin embargo lo anterior puede ser solucionado al definir un sistema de ecuaciones de mayor orden (Gugushvil, 2012).

En línea con el enfoque Bayesiano y los métodos de discretización anteriormente explicados, Stuart (2010) hace hincapié en evitar en lo posible la discretización pues menciona que uno de los retos que afronta la matemática científica, es la de dar un marco matemático y algorítmico coherente que permita combinar modelos matemáticos complejos con los conjuntos de datos experimentales provenientes de la aplicación científica, lo que ayudaría a tener una mejor comprensión de los fenómenos. Stuart (2010) utiliza un enfoque Bayesiano aplicado a espacios funcionales con el fin de obtener una medida (función utilizada para cuantificar el tamaño de un subconjunto), en el sentido Bayesiano, construir una fórmula para "medir" la distribución posterior a partir de datos e información a priori.

El enfoque realizado por Stuart (2010) requiere de una nueva notación, que describiremos como sigue. Dado un sistema de la forma

$$\mathbf{y} = \mathbf{x}(\boldsymbol{\theta}) \tag{2.7}$$

De donde se desea obtener una solución para $\boldsymbol{\theta} \in X$ dado $\mathbf{y} \in Y$, donde X, Y, son espacios de Banach. Definamos a **x** como el operador observación y llamemos a **y** los datos. El problema de mínimos cuadrados (2.2) puede ser entonces traducido a este contexto (2.8), ahora para alguna norma $\|\cdot\|_Y$ en el espacio Y.

$$arg\,min_{\boldsymbol{\theta}\in X}\frac{1}{2}\left\|\mathbf{y}-\mathbf{x}\left(\boldsymbol{\theta}\right)\right\|_{Y}^{2} \tag{2.8}$$

Sin embargo, este enfoque también presenta problemas, pues pueden existir sucesiones de parámetros, $\boldsymbol{\theta}^{(n)}$, que minimicen (2.8) pero que no converjan a un límite en X, o poseer múltiples mínimos (como el enfoque tradicional). Tales problemas pueden ser atacados de alguna manera al resolver un problema de minimización regularizado para un espacio de Banach $(E,\left\|\cdot\right\|_{E})$ contenido en X y un punto m_0, es decir,

$$arg\,min_{u\in E}\left(\frac{1}{2}\left\|\mathbf{y}-\mathbf{x}\left(\boldsymbol{\theta}\right)\right\|_{Y}^{2}+\frac{1}{2}\left\|u-m_0\right\|_{E}^{2}\right)$$

La elección de $\left\|\cdot\right\|_{E}$, $\left\|\cdot\right\|_{Y}$ y el punto m_0 son de alguna manera arbitrarios si no se realiza ninguna suposición adicional al modelaje. El enfoque Bayesiano, bajo esta perspectiva, lleva a la búsqueda de una medida de probabilidad μ^{y} en X, que contenga información sobre la probabilidad relativa de diferentes estados de $\boldsymbol{\theta}$, dado los datos $\mathbf{y}$. La derivación de la medida de probabilidad μ^{y} forzará la confrontación de varios modelos y supuestos matemáticos que guiarán a la elección de las normas $\left\|\cdot\right\|_{E}$, $\left\|\cdot\right\|_{Y}$ y el punto m_0. Dado que nuestro principal interés es el de tratar datos experimentales, reescribamos el sistema (2.7), ahora sujeto a una fuente de perturbación (2.9), donde $\mathbf{e}$ es una variable aleatoria desconocida de media cero. Describimos nuestro conocimiento a priori sobre $\boldsymbol{\theta}$, en términos de una medida de probabilidad μ_0, para posteriormente calcular mediante la fórmula de Bayes la medida de probabilidad posterior μ^{y}, para $\boldsymbol{\theta}$ dado $\mathbf{y}$.

$$\mathbf{y}=\mathbf{x}\left(\boldsymbol{\theta}\right)+\mathbf{e} \tag{2.9}$$

Ejemplificando en este contexto, considere a $\boldsymbol{\theta}\in\mathbb{R}^{n}$, $\mathbf{y}\in\mathbb{R}^{q}$ y sean π_0 y π^{y} las funciones de distribución de probabilidad de las medidas μ_0 y μ^{y}. Además, sea $\mathbf{e}\in\mathbb{R}^{q}$ una variable aleatoria con densidad ρ, entonces la probabilidad de y dado u tiene densidad dada por:

$$\rho\left(\mathbf{y}|\boldsymbol{\theta}\right):=\rho\left(\mathbf{y}-\mathbf{x}\left(\boldsymbol{\theta}\right)\right)$$

La cual representa la verosimilitud de los datos. De la mano de la fórmula de Bayes (2.4) y (2.5), se sigue que la distribución posterior π^{y} de la medida μ^{y} está dada por

$$\pi^{y}\left(\boldsymbol{\theta}\right)=\frac{\rho\left(\mathbf{y}-\mathbf{x}\left(\boldsymbol{\theta}\right)\right)\pi_0\left(\boldsymbol{\theta}\right)}{\int_{\mathbb{R}^{n}}\rho\left(\mathbf{y}-\mathbf{x}\left(\boldsymbol{\theta}\right)\right)\pi_0\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta}}$$

Dado que $\int_{\mathbb{R}^{n}}\rho\left(\mathbf{y}-\mathbf{x}\left(\boldsymbol{\theta}\right)\right)\pi_0\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta}$ sirve como un término normalizador, podemos expresar a $\pi^{y}\left(\boldsymbol{\theta}\right)$ como

$$\pi^y(\boldsymbol{\theta}) \propto \rho(\mathbf{y} - \mathbf{x}(\boldsymbol{\theta}))\,\pi_0(\boldsymbol{\theta})$$

De manera abstracta, la ecuación anterior expresa que la medida posterior μ^y (con densidad π^y) y la medida a priori μ_0 (con densidad π_0) están relacionadas por medio de la derivada Radon-Nikodym

$$\frac{d\mu^y}{d\mu_0}(\boldsymbol{\theta}) \propto \rho(\mathbf{y} - \mathbf{x}(\boldsymbol{\theta}))$$

Sin pérdida de generalidad, al ser ρ una densidad y por lo tanto no negativa, la expresión anterior en su lado derecho puede representarse por medio de la exponencial de un potencial negativo $\Phi(\boldsymbol{\theta}; \mathbf{y})$

$$\frac{d\mu^y}{d\mu_0}(\boldsymbol{\theta}) \propto \exp(-\Phi(\boldsymbol{\theta}; \mathbf{y}))$$

Forma que generaliza a una situación donde X y posiblemente Y sean infinito - dimensionales. En el caso en que X y Y sean finito - dimensionales, las perturbaciones en (2.9) son aditivas y Gaussianas y la medida a priori es Gaussiana, la medida posterior tendrá densidad π^y dada por

$$\pi^y(\boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2}\|\mathbf{y} - \mathbf{x}(\boldsymbol{\theta})\|_Y^2 - \frac{1}{2}\|u - m_0\|_E^2\right)$$

2.4. Manejo de incertidumbre

Como se ha visto, existe un gran número de técnicas de estimación, sin embargo, es necesario reconocer la incertidumbre propia del contexto sobre los parámetros del modelo; así como de las predicciones producto de la estimación. Vanlier ***et al***. (2013), considera algunas propiedades en el proceso de estimación que permiten cuantificar en cierta forma la incertidumbre en los parámetros y medir la confianza en el modelo.

Sensibilidad

El análisis de sensibilidad consiste en la perturbación de los parámetros del modelo y la observación del efecto sobre la predicción del mismo. Por medio de este análisis se puede identificar si el sistema es robusto en el sentido de que es capaz de operar de manera confiable cuando sus parámetros físicos y bioquímicos varían dentro de sus rangos esperados.

Parámetros y componentes sensibles a variaciones pueden introducir fragilidad en el sistema (Riel, 2006). Existen dos vertientes de tal análisis, uno local y uno global. El primero consta de la perturbación de un parámetro a la vez, mientras el segundo resulta de la perturbación simultánea de múltiples parámetros. La interpretación de ambos enfoques debe ser cuidadosa, especialmente al tenerse parámetros altamente inciertos, donde el análisis provee de una valoración de lo que ocurre al variar los parámetros y no así el de dar una idea del efecto de la incertidumbre parametral. Técnicas, como el Filtro Monte Carlo, son utilizadas para clasificar simulaciones de los parámetros como aceptadas o no; en este sentido se crean distribuciones de los parámetros aceptados (y no aceptados) para posteriormente calcular una distancia (distancia Kolmogorov-Smirnov) entre tales densidades como una forma de observar qué tan fuerte la aceptación y rechazo se correlaciona con el parámetro específico (Vanlier *et al*., 2013).

Identificabilidad

La modelación y simulación permite generar predicciones y nuevas hipótesis, sin embargo, la calidad de tales interpretaciones y de sus resultados dependerá plenamente de la calidad del modelo. En este sentido, la estimación de parámetros se vuelve un paso crítico en el proceso de modelaje. A pesar de la disponibilidad y calidad de los datos experimentales, el proceso de estimación permanece aún como una dificultad enorme desde el punto de vista matemático y computacional. Tales dificultades se originan, generalmente, de la carencia de identificabilidad del sistema, es decir, de la dificultad o imposibilidad de asignar valores únicos a los parámetros desconocidos o de determinar intervalos finitos de confianza (Chis *et al*., 2011). Los métodos tradicionales de estimación generan un juego de parámetros que optimizan de cierta manera el ajuste del modelo a los datos, sin embargo, resulta imprudente realizar conclusiones con base a un único ajuste. En este sentido es necesario tratar de evaluar el comportamiento del modelo ante todos aquellos parámetros que cumplen con describir los datos experimentales con un cierto grado de fiabilidad.

El análisis resultante lleva el nombre de análisis de identificabilidad. Un modelo es identificable si se caracteriza por poseer intervalos de confianza finitos. Su ausencia, la no identificabilidad, puede ser dividida en dos clases: no-identificabilidad estructural, propia de los parámetros del modelo, independiente de los datos y se relaciona a una parametrización redundante y está caracterizada por un conjunto de parámetros que puede se variado sin perturbar la solución significativamente. Los intervalos de confianza de un parámetro con este tipo de no identificabilidad resultan ser infinitos (Schaber y Klipp, 2011). El segundo tipo de problema de no identificabilidad se denomina práctico, que resulta de la falta de información en los datos y se manifiesta como una verosimilitud plana, y se caracteriza por que los parámetros poseen intervalos de confianza infinitos respecto a un cierto nivel de confianza, es decir, define

a un cierto parámetro como arbitrario en ciertas regiones. En el enfoque Bayesiano ataca este problema al incorporar, distribuciones de probabilidad a priori de tipo informativas o aportando datos adicionales, lo que se denomina identificabilidad débil (Vanlier *et al*., 2013; Schaber y Klipp, 2011).

Asintoticidad

Otra propiedad de interés en la estimación de parámetros es la de poder determinar la facilidad con la que se pueden estimar los parámetros. Si el modelo es identificable y bajo ciertas propiedades de regularidad, el estimador máximo verosímil tiende a una distribución Gaussiana para un número lo suficientemente grande de datos, como consecuencia se pueden estimar intervalos de confianza aproximados utilizando la distribución t de Student. Tales intervalos no siempre resultan ser adecuados pues como se ha mencionado, dada la dependencia en la identificabilidad del modelo, no siempre se cuenta con suficientes datos que puedan respaldar este supuesto (Vanlier *et al*., 2013).

Verosimilitud Perfil

La construcción de la función de verosimilitud puede enfrentarse a topologías complejas y por ende resultar en regiones con múltiples subóptimos. En este sentido se puede utilizar la verosimilitud perfil para obtener intervalos de confianza más confiables. Se basa en el trazado de una trayectoria óptima sobre la verosimilitud, o en el caso de la MAP una función de densidad de probabilidad óptima. Cada perfil se inicia en el mejor ajuste de los parámetros donde posteriormente uno de los parámetros es seleccionado para ser perfilado y posteriormente cambiado y al mismo tiempo re-optimizar los parámetros restantes. Tal procedimiento es iterado hasta que el ajuste se vuelva inaceptable o cumpla cierto umbral (Vanlier *et al*., 2013).

Métodos Basados en Muestreo

El manejo de la incertidumbre, como se ha observado, requiere de ciertos supuestos que muchas de las veces no se cumplen. Una alternativa para el manejo de incertidumbre son los denominados métodos basados en muestreo. Tales procedimientos nuevamente divergen dependiendo del enfoque que se esté utilizando, frecuentista o Bayesiano, que ya se han explicado con anterioridad. En el primero se utiliza una técnica denominada bootstrap, cuyo objetivo es el de muestrear de las réplicas experimentales disponibles y realizar una estimación para cada conjunto muestreado, para posteriormente conformar una muestra de parámetros estimados y por tanto una posible distribución de los mismos que da cierto entendimiento sobre

la incertidumbre de los parámetros. En esta técnica se desprenden dos clases: el boostrap paramétrico y el no paramétrico.

Bootstrap

El boostrap paramétrico se basa en ajustar un modelo a los datos y obtener muestras de la distribución del error parametrizado. En este sentido se está suponiendo que el modelo utilizado para generar las muestras es lo suficientemente confiable como para ser equiparado al proceso que genera los datos experimentales. Un procedimiento alternativo para obtener tales muestras corresponde al uso de procesos Gaussianos (PG), que corresponde a un conjunto de variables definidas por una media y una función de covarianza que describe cierta relación entre los elementos del conjunto. Mediante la utilización del teorema de Bayes (2.5) los parámetros del modelo son actualizados a una distribución posterior; de la cual las muestras pueden ser obtenidas y posteriormente se les puede suministrar ruido para ser usadas como muestras bootstrap.

La segunda clase, el bootstrap no paramétrico, se refiere al muestreo con reemplazo de las réplicas experimentales disponibles. La idea principal es que la variabilidad de las estimaciones alrededor de los valores reales es imitada por las estimaciones basadas en bootstrap sobre las réplicas alrededor del estimador. Cabe señalar que este enfoque requiere de un suficiente número de réplicas para capturar adecuadamente la variabilidad en la muestra. Tras el uso de las técnicas de bootstrap se pueden obtener intervalos de confianza, sin embargo, tienden a subestimar la verdadera región de confianza y por tanto requieren ser corregidos (Vanlier *et al*., 2013).

Mientras los métodos deterministas se centran en la minimización y cálculo de cotas para intervalos de confianza, la alternativa Bayesiana utiliza integración intensiva. En este sentido, este enfoque requiere de muestreo sobre las densidades de probabilidad inmiscuidas, para así determinar cotas que contengan un intervalo al $(1-\alpha)\,100$ % de la densidad objetivo, sin embargo, integración numérica simple no resulta ser una opción, dado el alto costo computacional requerido, por lo que alternativas como el muestreo MCMC son utilizadas.

Técnicas Markov Chain Monte Carlo (MCMC)

En la práctica, un muestreo uniforme sobre el espacio de parámetros resulta ser extremadamente ineficiente debido a que las verosimilitudes obtenidas de cada muestra resultan tener regiones de baja probabilidad, lo que se traduce en la presencia de óptimos locales. Otra alternativa es el muestreo por importancia, que consiste en muestrear de acuerdo a cierta función de densidad de probabilidad (g), en vez de la función de densidad objetivo. Es claro

entonces que para obtener buenos resultados, la distribución g debe ser tan cercana a la distribución objetivo como se pueda, sin embargo para modelos no lineales no es trivial encontrar tal función g. Técnicas para evitar problemas como los que presenta el muestreo por importancia son las denominadas técnicas Markov Chain Monte Carlo (MCMC) (Vanlier *et al*., 2013).

Los métodos MCMC son técnicas de simulación de procesos estocásticos que poseen distribuciones de probabilidad conocidas hasta alguna cierta constante de proporcionalidad. Pueden ser usadas para simular una amplia variedad de variables aleatorias y procesos estocásticos y son de gran utilidad en la inferencia estadística frecuentista, Bayesiana y de verosimilitud (Geyer, 1992). Trabajan muestreando directamente de la distribución de densidad objetivo, denominada distribución posterior. Con el fin de asegurar la convergencia del muestreo, se requiere que la distribución posterior sea débilmente identificable, es decir, que eligiendo una a priori lo suficientemente informativa, se pueden rebasar regiones de baja probabilidad en la función de densidad de probabilidad posterior. Algoritmos dentro de este tipo de métodos de muestreo son el Metropolis-Hastings y el population-based MCMC, que forman parte del estudio de la presente investigación.

Método Metropolis-Hastings

Basado en la aplicación de reglas de salto para el muestreo de un cierto $\boldsymbol{\theta}^*$ de una distribución de salto (jumping distribution) en un tiempo t, con el objetivo de muestrear de la distribución posterior, el muestreo es realizado de forma secuencial de tal forma que la distribución de las muestras es dependiente sólo del último valor obtenido; conformando así, una cadena de Markov; de ahí su nombre. Una cadena de Markov es una sucesión de variables $\boldsymbol{\theta}^1$, $\boldsymbol{\theta}^2$, ..., con la propiedad de que para cualquier valor de t, la distribución de $\boldsymbol{\theta}^t$ sólo depende del valor más reciente, es decir $\boldsymbol{\theta}^{t-1}$. El punto esencial es que la cadena es elegida de tal forma que la distribución converge a $p\left(\boldsymbol{\theta}|\mathbf{y}\right)$ (Congdon, 2006).

La metodología se resume de Gelman *et al*., (2004) como:

1. Se toma un punto inicial $\boldsymbol{\theta}^0$, tal que $p\left(\boldsymbol{\theta}^0|\mathbf{y}\right) > 0$ de una distribución inicial $p_0\left(\boldsymbol{\theta}\right)$.

2. Para $t = 1, 2, ...$:

a) Muestrear un $\boldsymbol{\theta}^*$ propuesto de una distribución de salto (o distribución propuesta) en una tiempo t, $J\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{t-1}\right)$.

b) Se calcula el cociente entre las densidades

$$r = \frac{p\left(\boldsymbol{\theta}^*|\mathbf{y}\right)/J_t\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{t-1}\right)}{p\left(\boldsymbol{\theta}^{t-1}|\mathbf{y}\right)/J_t\left(\boldsymbol{\theta}^{t-1}|\boldsymbol{\theta}^*\right)}$$

c) Se define

$$\boldsymbol{\theta}^t = \begin{cases} \boldsymbol{\theta}^* & \text{con probabilidad } min\left(r,\, 1\right) \\ \boldsymbol{\theta}^{t-1} & \text{en otro caso} \end{cases}$$

Dado el valor actual de $\boldsymbol{\theta}^{t-1}$, la distribución de transición $T_t\left(\boldsymbol{\theta}^t|\boldsymbol{\theta}^{t-1}\right)$ de la cadena de Markov es una mezcla de puntos de masa en $\boldsymbol{\theta}^t = \boldsymbol{\theta}^{t-1}$, y una versión ponderada de la distribución de salto $J_t\left(\boldsymbol{\theta}^t|\boldsymbol{\theta}^{t-1}\right)$, que se ajusta al cociente de aceptación. El algoritmo requiere la habilidad de computar la razón r, para todo par $(\boldsymbol{\theta},\, \boldsymbol{\theta}^*)$ y tomar $\boldsymbol{\theta}$ de la distribución de salto para todo $\boldsymbol{\theta}$ y t. El objetivo es muestrear de la distribución objetivo mediante $J\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}\right) = p\left(\boldsymbol{\theta}^*|\mathbf{y}\right)$ para todo t (Gelman *et al*., 2004).

Tal ejecución permite implementar múltiples cadenas de forma simultánea para monitorear la convergencia del muestreador a la distribución posterior (Gelman *et al*., 2004). Uno de los inconvenientes es que el uso de a prioris dispersas o alejadas de la distribución objetivo produce tasas de aceptación bajas. Tasas de aceptación muy altas generan una alta autocorrelación en la muestra debido a la restricción de la cadena a un espacio determinado, mientras que tasas de aceptación bajas tienden a provocar el mismo problema al fijarse en valores determinados (Congdon, 2006). Tales tasas de aceptación pueden ser mejoradas al concentrar el muestreo cerca del estado actual de la cadena, pero ello se traduce en una lenta e ineficiente búsqueda en el espacio de parámetros. El problema de muestreo local sumado a la autocorrelación producida por las sucesivas observaciones se traduce en una mezcla pobre sobre el espacio de parámetros.

La eficiencia de este algoritmo se ve incrementada al usar una distribución a priori tan cercana como sea posible de la distribución objetivo, por el hecho de que distribuciones planas o muy dispersas, incluyen regiones multimodales cayendo así en el problema del Local-Trap (Congdon, 2006; Faming *et al*., 2010). Este tipo de método en situaciones de multimodalidad (como en el caso de modelos oscilatorios) puede converger prematuramente a un modo subóptimo dejando sin explorar regiones con mejores soluciones.

Método population-based MCMC Jasra (2007)

Una alternativa para evitar el problema del Local-Trap que sufre el algoritmo Metropolis-Hastings es el population-based MCMC. Este método utiliza técnicas como el Parallel Tempering (PT) y el intercambio de información para resolver problemas de modas locales. Distribuciones multimodales pueden resultar un problema para las técnicas de simulación MCMC (Vyshemirsky y Girolami, 2008). Dado que el objetivo es muestrear toda la distribución posterior, esto implica muestrear cada moda con una probabilidad significativa. Sin embargo, resulta

fácil a las cadenas de Markov permanecer en una vecindad de tales modas por un largo tiempo. Lo anterior ocurre cuando las modas se encuentran separadas por regiones de baja densidad. Resultando así que saltos de tipo Metropolis-Hastings entre tales regiones sean rechazados.

La técnica PT es una estrategia para mejorar la eficiencia de barrido del espacio parametral en la simulación MCMC. Con la distribución $p(\boldsymbol{\theta}|\mathbf{y})$ como objetivo, esta técnica trabaja un conjunto de $K+1$ simulaciones MCMC, cada una de ellas con su propia distribución estacionaria $p_k(\boldsymbol{\theta}|\mathbf{y})$, $k = 1, 2, ..., K$, donde $p_0(\boldsymbol{\theta}|\mathbf{y}) = p(\boldsymbol{\theta}|\mathbf{y})$, y $p_1, ..., p_k$ son distribuciones con la misma forma básica pero con probabilidades altas para barrer el espacio a través de las distintas modas.

- La sucesión de densidades $\{p_k(\boldsymbol{\theta}|\mathbf{y})\}$ será seleccionada de tal forma que todas estén relacionadas, y en general, más fáciles de simular que $p(\boldsymbol{\theta}|\mathbf{y})$.
- El uso de una población de muestreadores permitirá movimientos más globales (que una cadena simple de tipo MCMC), construidas de tal forma que resulten en un mezclado más rápido de los algoritmos MCMC (Jasra *et al*., 2005).

Es decir, aproxima la distribución posterior de $\boldsymbol{\theta}$ a través de una sucesión de $k = 1, ..., K$ aproximaciones $p_k(\boldsymbol{\theta}|\mathbf{y}) \approx p(\boldsymbol{\theta}|\mathbf{y})$ definidas por un gradiente de amortiguado $0 \leq \gamma_1 < \cdots < \gamma_k = 1$, donde $p_k(\boldsymbol{\theta}|\mathbf{y}) \propto (p(\mathbf{y}|\boldsymbol{\theta}))^{\gamma_k} p(\boldsymbol{\theta})$ (Campbell y Steele, 2012).

Problemas de lentitud de mezclado producidos por modas locales se ven resueltos al introducir kernels de transición con distribuciones estacionarias con diversos grados de dispersión, lo anterior produce poblaciones de muestras que permiten movimientos más globales que una cadena simple de tipo Metropolis-Hastings. Esto último resulta favorable cuando se tienen modelos de alta complejidad o modelos de tipo oscilatorio (Vyshemirsky & Girolami, 2008).

Otra alternativa para muestrear de la distribución posterior son las técnicas de muestreo Monte Carlo Secuencial, en los cuales a diferencia de tenerse múltiples cadenas en interacción, propagan conjuntos de parámetros, denominados partículas, a través de series de distribuciones intermedias, sin embargo tales distribuciones con el tiempo se vuelven mas difíciles de muestrear. Alternativas donde se evita por completo el muestreo y uso de técnicas de tipo MCMC han sido desarrolladas con el fin de mitigar el alto costo y tiempo computacional requerido por las anteriores técnicas (Vanlier *et al*., 2013).

Simulación libre de MCMC

Métodos que evitan el muestreo de las funciones de probabilidad de los modelos, en lugar de resolver el sistema de ecuaciones ordinarias, estiman las derivadas de los estados del

sistema permitiendo trabajar con el sistema de ecuaciones directamente. Tales técnicas hacen uso de aproximaciones de las dinámicas observadas por medio de splines para posteriormente realizar la optimización. Esta aproximación, sin embargo, requiere que todos los estados del sistema hayan sido medidos además de requerir de un parámetro que regule el nivel con que la aproximación por funciones base se ajuste a los datos. Este enfoque fue generalizado por Ramsay *et al*. (2007), donde el ajuste por splines y el parámetro de ajuste a los datos (fidelidad) fueron combinados en un mismo paso de estimación (Vanlier *et al*., 2013).

La estimación de parámetros y la cuantificación de la incertidumbre, hasta ahora mencionados, han generado un panorama de algunos de los métodos disponibles para la estimación de parámetros. Cada uno provee de formas particulares de tratar la incertidumbre muestral así como de pros y contras en su aplicación. El proceso de modelaje proporciona las herramientas para realizar predicciones tanto cualitativas como cuantitativas. Tras obtener un modelo parametrizado, se pueden realizar predicciones cuantitativas sobre el comportamiento de componentes del sistema que aún no se han medido, así como predicciones cualitativas de posibles procesos que rigen el comportamiento observado o el de análisis de interacciones complejas antes de pruebas experimentales. Tras haber podido verificar nuestras predicciones de manera experimental ganamos confianza en nuestro modelo al afirmar que puede capturar todos aquellos procesos de importancia que explican nuestro fenómeno.

Al comparar varios modelos candidatos y realizando algún análisis de discriminación sobre los modelos, se pueden hacer predicciones acerca de los procesos aún desconocidos. Es evidente que es más fácil cambiar un parámetro en un modelo matemático que el de sintonizar una característica bioquímica en vivo. Por lo tanto, los modelos bien definidos son útiles para predecir resultados mucho antes de realizar un experimento. Es así que un buen modelo se traduce como una prueba rápida y de bajo costo del proceso experimental (Schaber y Klipp, 2011).

3. Comparación de Métodos de Estimación Bayesiana en Sistemas Estables y Cíclicos

Introducción

Como cualquier otro método de estimación, los métodos de muestreo de tipo MCMC se enfrentan a ventajas y desventajas surgidos de la forma en que fueron construidos. Con el fin de observar el desempeño y las eficiencias de los métodos de estimación Bayesianos aquí estudiados, se procedió a la comparación de los mismos. Para ello se procedió a la estimación de parámetros de un par de sistemas de ecuaciones diferenciales ordinarias no lineales del tipo Presa-Depredador (Lotka-Volterra). Tales sistemas describen la competencia por los recursos entre dos especies bajo distintos escenarios y de distinta complejidad.

El fenómeno presa-depredador se caracteriza por la interacción entre especies competidoras de tal manera que la ausencia de una provoca la disminución o desaparición de la otra. Tal interacción genera un ciclo de aumento y disminución poblacional de manera periódica, sin embargo, dado que el ambiente también determina la capacidad para sostener una especie, existe una variante del fenómeno donde las poblaciones están, además, sujetas a una cierta capacidad ambiental (Spiegel, 1983). Para nuestros fines, y sin pérdida de generalidad, consideraremos dos especies, conejos y zorros, que hacen el rol de presas y depredadores, respectivamente.

3.1. Modelos

Denominaremos a la población de conejos por R (Rabbits) y la de zorros por F (Foxes), por su traducción a inglés. Se consideran dos modelos, ambos con cuatro parámetros y sujetos a cuatro hipótesis comunes más una quinta para el caso específico del segundo modelo, como se menciona en Blanchard *et al*. (1999):

1. Si no hay zorros presentes, los conejos se reproducen a una tasa proporcional a su población y no afecta la sobrepoblación.

2. Los zorros se comen a los conejos y la proporción a la que los conejos son devorados es proporcional a la tasa a la que los zorros y conejos interactúan.

3. Sin conejos que comer, la población de zorros declina a una razón proporcional a ella misma.

4. La tasa de nacimientos de los zorros va en proporción al número de conejos comidos por zorros que, por la segunda hipótesis, es proporcional a la razón a la que los zorros y conejos interactúan.

5. Supone que en ausencia de depredadores la población obedece a un modelo logístico de crecimiento en vez de uno exponencial. En este caso la capacidad del entorno para sustentar a la población de presas está dado en el modelo (3.2) por la expresión $1 - (R(t)/2)$, con límite de capacidad 2.

Los parámetros son cuatro y se describen como sigue:

a: Razón de crecimiento de conejos.

b: Número de interacciones conejos - zorros en las que el conejo es devorado.

c: Razón de muertes de zorros.

d: Beneficio a la población de zorros de un conejo devorado.

Bajo estas hipótesis se desprenden dos modelos: el primero modelo (3.1), corresponde a una dinámica oscilatoria (Figura 3.1), mientras el segundo modelo (3.2), corresponde a una dinámica que se estabiliza en el tiempo (Figura 3.1).

$$\begin{aligned} \frac{d}{dt}R(t) &= a \cdot R(t) - b \cdot F(t) R(t) \\ \frac{d}{dt}F(t) &= -c \cdot F(t) + d \cdot F(t) R(t) \end{aligned} \tag{3.1}$$

$$\begin{aligned} \frac{d}{dt}R(t) &= a \cdot R(t) [1 - R(t)/2] - b \cdot F(t) R(t) \\ \frac{d}{dt}F(t) &= -c \cdot F(t) + d \cdot F(t) R(t) \end{aligned} \tag{3.2}$$

La Figura 3.1 muestra las soluciones estimadas del par de modelos para las condiciones iniciales al tiempo cero, $R(0) = 4$ y $F(0) = 2$; y con parámetros $\boldsymbol{\theta}$ =($a_0 = 2$, $b_0 = 1.2$, $c_0 = 1$, $d_0 = 0.9$). De tales soluciones se observa que el modelo (3.1) presentan oscilaciones en su dinámica mientras que las trayectorias del modelo (3.2) tienden a estabilizarse con el tiempo.

Estimación

Para el proceso de estimación se simularon datos en un intervalo de tiempo $[0, 15]$ dividido en 300 puntos por medio de métodos numéricos. Posteriormente se añadió un ruido aditivo e con distribución $N(0, \sigma^2)$ y $\sigma = 0.5$. El proceso de estimación requiere de distribuciones a priori de los parámetros, suponiendo desconocimiento de los mismos se dieron a priori uniformes $U[0, 3]$ para cada uno de los 4 parámetros. El software se corrió hasta alcanzar un cierto límite de iteraciones, donde la literatura sugiere 200,000 iteraciones como límite de "*quemado*" (*Burn-In*). Finalmente se obtuvo una muestra de las distribuciones posteriores de los parámetros de tamaño 6,000 para cada uno de ellos, utilizando un Thinning muestral de 5 (iteraciones entre muestreo).

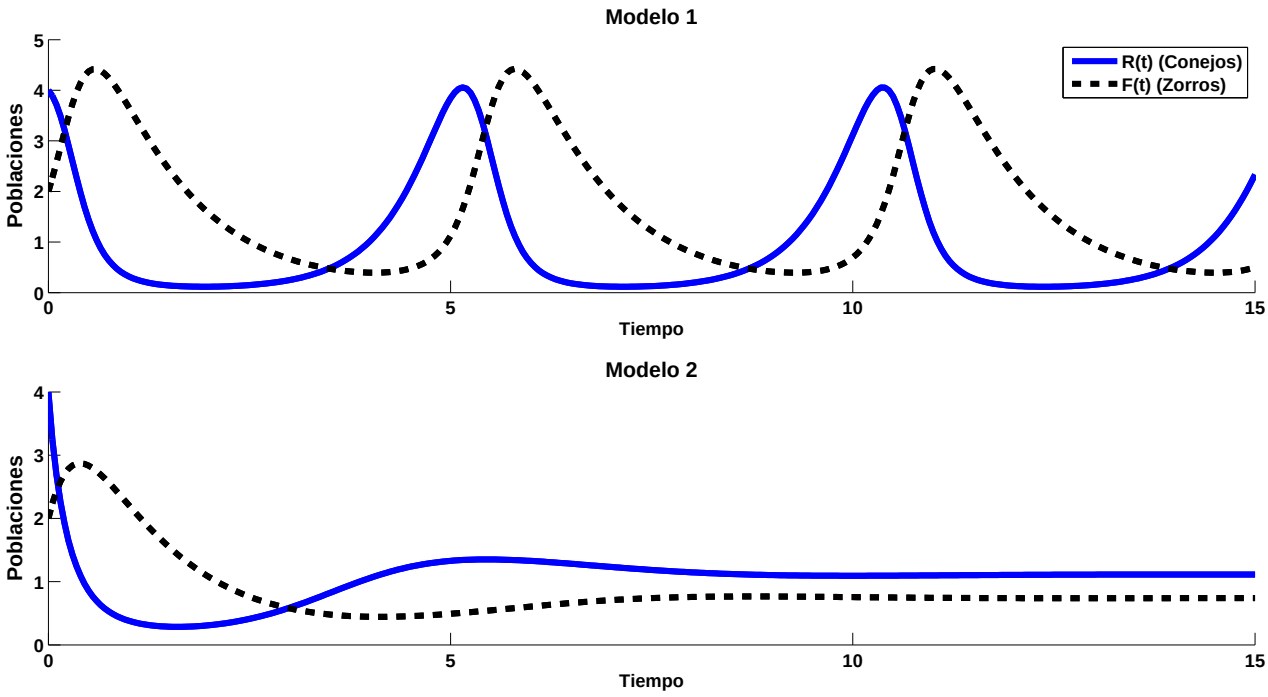

Figura 3.1: Por filas las soluciones para el sistema presa depredador para los modelos (3.1) y (3.2), respectivamente.

3.2. Algoritmos Metropolis-Hastings y population-based MCMC

Modelo (3.1)

La Figura 3.2 en sus dos columnas muestra las distribuciones posteriores de los parámetros del sistema presa-depredador para el modelo (3.1), estimadas mediante los métodos Metropolis-Hastings y population-based MCMC, respectivamente. En la primera columna (método Metropolis-Hastings) se observan distribuciones multimodales con regiones de alta densidad en los extremos de la distribución, particularmente alrededor del cero. Tal fenómeno es más evidente en los parámetros c y d (filas 3 y 4, respectivamente), situación que se repite en la segunda columna (método population-based MCMC). La existencia de tales regiones de probabilidad en los extremos de las distribuciones, se traduce en la estimación de intervalos de probabilidad bastante extensos y por lo tanto, poco informativos. En general, los intervalos de probabilidad para los parámetros a y b (filas 1 y 2, respectivamente), resultan ser más reducidos en el caso de la estimación por medio del método population-based MCMC, y estas distribuciones son más concentradas que en el caso del método Metropolis-Hastings. Respecto a las estimación del parámetro (línea roja cortada), se encuentra cercana al valor real del parámetro (línea negra punteada), cabe señalar que tal estimación corresponde a la moda de la distribución posterior respectiva.

Siguiendo con este análisis comparativo y con el fin de exhibir el problema del Local-Trap, se realizó una segunda estimación (que denominaremos réplica) de parámetros para ejemplificar los problemas que presenta el método Metropolis-Hastings ante sistemas complejos como lo es el modelo (3.1). La nueva estimación se realiza bajo los mismos parámetros y condiciones iniciales de los usados con anterioridad. Lo que refleja únicamente el carácter aleatorio de la estimación ante el problema ya mencionado.

La Figura 3.3 presenta nuevamente las distribuciones posteriores de los parámetros del modelo (3.1) en una segunda estimación, obtenidas a partir de los métodos Metropolis-Hastings y population-based MCMC, respectivamente. Se observan distribuciones similares al de la primera estimación (Figura 3.2), sin embargo es claro que para los parámetros c y d, estimados por medio del método Metropolis-Hastings (Figura 3.3, columna 1, filas 3 y 4), el parámetro estimado (línea vertical roja cortada) resulta estar muy alejado del valor real del parámetro (línea negra vertical punteada). Su localización es ahora en una región con una alta densidad dentro de la distribución bimodal, de hecho, el intervalo de probabilidad contiene valores negativos, lo que resulta absurdo al ser los parámetros estrictamente positivos. Si bien en la segunda columna, correspondiente al método population-based MCMC, las distribuciones son bimodales para los parámetros c y d (filas 3 y 4), el método arroja un estimador muy cercano al valor real del parámetro (totalmente contrario al primer método, para estos parámetros). En el caso de los parámetros a y b (filas 1 y 2, Figura 3.3) el método population-based MCMC (segunda columna)

recupera distribuciones de probabilidad concentradas alrededor del valor real del parámetro, situación que no ocurre con el método Metropolis-Hastings (primer columna), que nuevamente posee distribuciones dispersas y con intervalos de probabilidad muy extensos.

Modelo (3.2)

Siguiendo con la estimación de parámetros, ahora se ha de considerar el modelo (3.2), que corresponde a una dinámica que se estabiliza en el tiempo y carece de las oscilaciones presentes en el modelo (3.1). La Figura 3.4 muestra las distribuciones posteriores para los parámetros (filas) a partir de los métodos Metropolis-Hastings y population-based MCMC (columnas). Se observa una marcada diferencia respecto a las distribuciones obtenidas para el modelo (3.1), puesto que ahora se tienen distribuciones unimodales y con intervalos de probabilidad más reducidos. Los estimadores dados por la moda de la distribución (línea vertical roja cortada) resultan ser muy similares en ambos métodos.

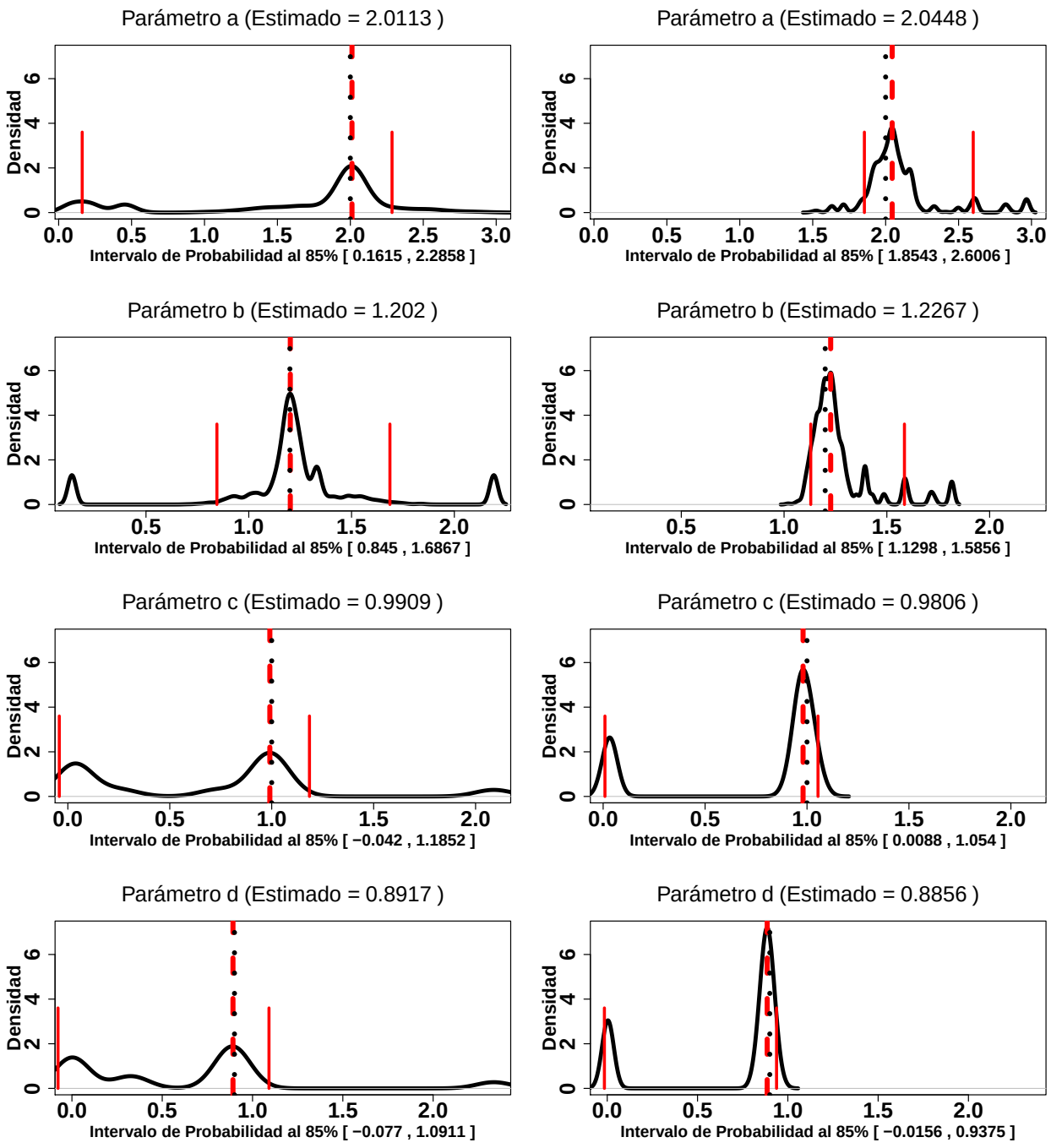

Figura 3.2: Distribución de densidad posterior para los parámetros estimados por los métodos Metropolis-Hastings y population-based MCMC (columnas) a partir de los datos del modelo (3.1). Por filas, la distribución posterior aproximada de los parámetros a, b, c y d, respectivamente. Se indica el valor real por la línea punteada negra y el estimado por la línea cortada roja. Las líneas verticales rojas y sólidas corresponden a los límites inferior y superior de los intervalos de probabilidad al 85 %.

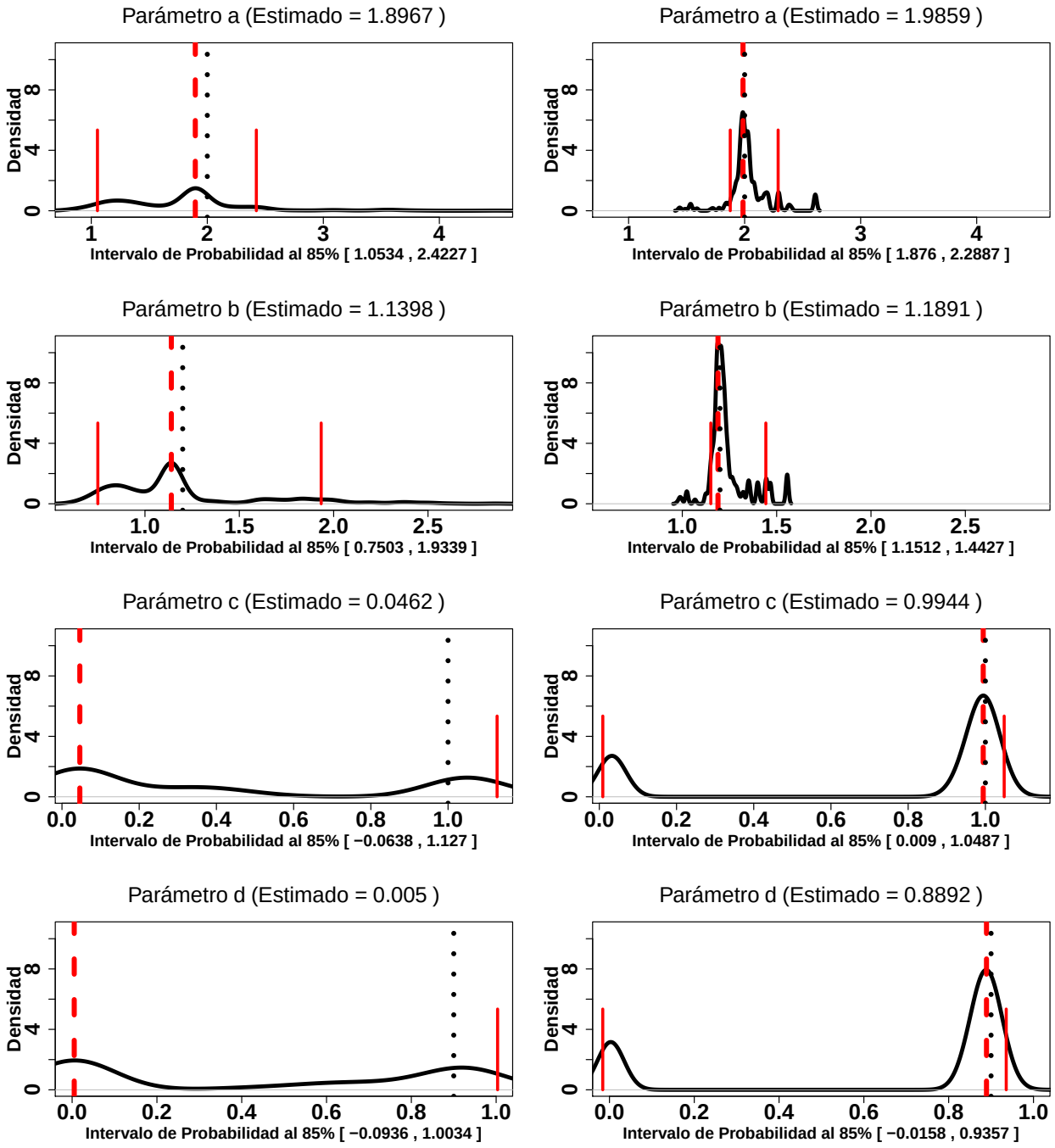

Figura 3.3: Distribución de densidad posterior para los parámetros estimados por los métodos Metropolis-Hastings y population-based MCMC (columnas) a partir de los datos del modelo (3.1) en una segunda estimación. Por filas, la distribución posterior aproximada de los parámetros a, b, c y d, respectivamente. Se indica el valor real por la línea punteada negra y el estimado por la línea cortada roja. Las líneas verticales rojas y sólidas corresponden a los límites inferior y superior de los intervalos de probabilidad al 85 %.

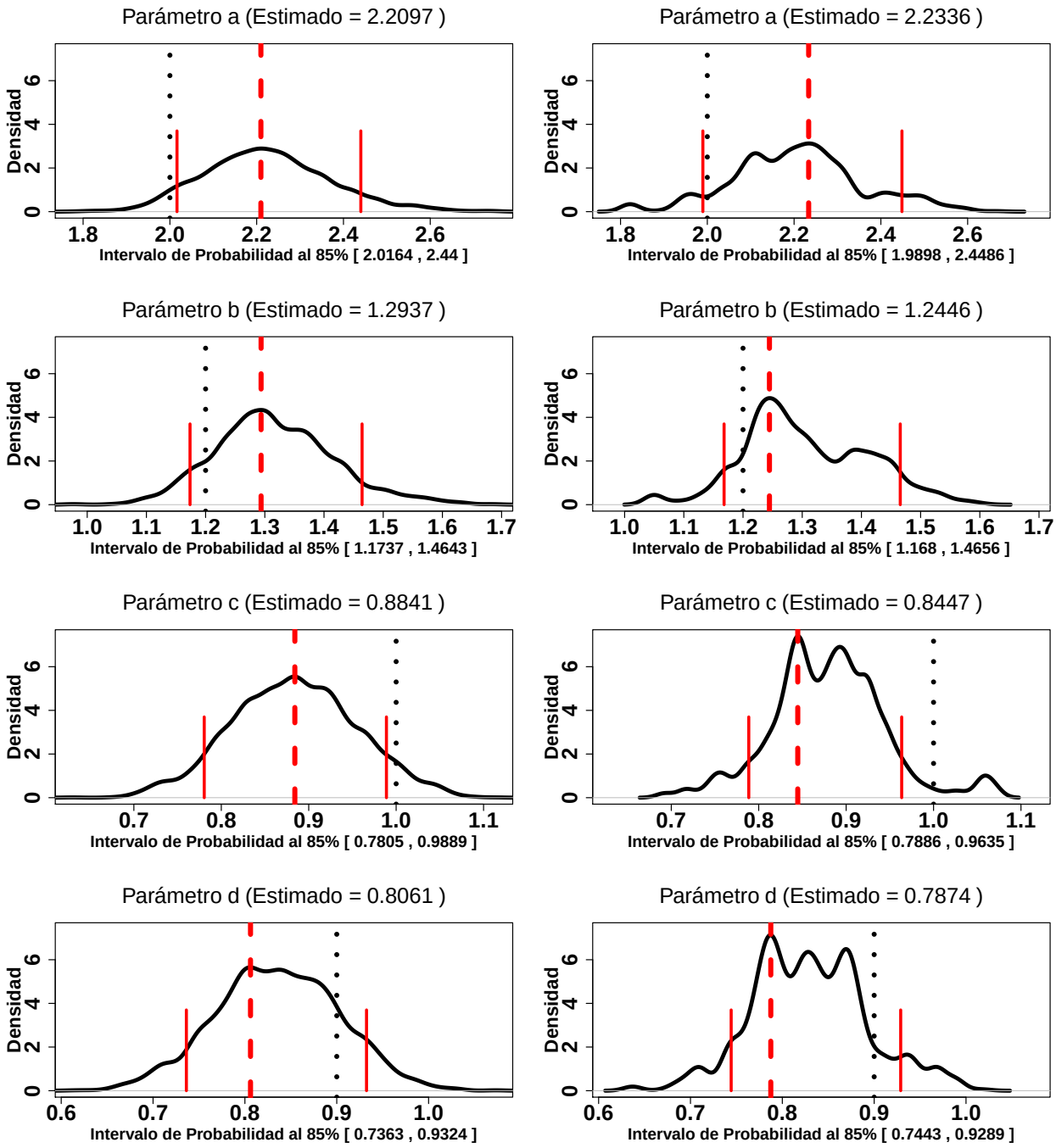

Figura 3.4: Distribución de densidad posterior para los parámetros estimados por los métodos Metropolis-Hastings y population-based MCMC (columnas) a partir de los datos del modelo (3.2). Por filas, la distribución posterior aproximada de los parámetros a, b, c y d, respectivamente. Se indica el valor real por la línea punteada negra y el estimado por la línea cortada roja. Las líneas verticales rojas y sólidas corresponden a los límites inferior y superior de los intervalos de probabilidad al 85 %.

3.3. Soluciones estimadas

Los Cuadros 3.1, 3.2 y 3.3, contienen los parámetros estimados por la moda de las distribuciones posteriores a partir de los métodos Metropolis-Hastings y population-based MCMC, correspondientes a las Figuras 3.2, 3.3 y 3.4, respectivamente.

Las Figuras 3.5 y 3.6 muestran por filas las soluciones versus los datos para las funciones $R(t)$ y $F(t)$, respectivamente, a partir de los parámetros de los Cuadros 3.1 y 3.2. Las soluciones estimadas por el método Metropolis-Hastings se muestran por medio de la línea magenta continua (MH) y las soluciones estimadas por medio del método population-based MCMC se muestran por la línea negra cortada (PB). Ambas estimaciones resultan ser muy similares en la primera réplica de estimación, sin embargo, al observar las soluciones estimadas (Figura 3.6, Cuadro 3.2) a partir de los parámetros obtenidos de la segunda estimación (Figura 3.3), se observa una completa pérdida de los datos por parte de la solución estimada por medio del método Metropolis-Hastings, debido al Local-Trap.

Con respecto al modelo (3.2), el Cuadro 3.3 y la Figura 3.7 muestran respectivamente los parámetros y soluciones estimadas. Las soluciones resultan ser bastante cercanas entre sí, adecuándose a los datos experimentales de manera razonable.

Cuadro 3.1: Parámetros estimados por los métodos Metropolis-Hastings y population-based MCMC a partir de la moda de la distribución posterior para el modelo (3.1).

Modelo (3.1)	a	b	c	d
Reales	**2**	**1.2**	**1**	**0.9**
Metropolis-Hastings	2.0013	1.202	0.9909	0.8917
Population-based MCMC	2.0448	1.2267	0.9806	0.8859

Cuadro 3.2: Parámetros estimados por los métodos Metropolis-Hastings y population-based MCMC a partir de la moda de la distribución posterior para el modelo (3.1) en una segunda réplica de estimación.

Modelo (3.1)	a	b	c	d
Reales	2	1.2	1	0.9
Metropolis-Hastings	1.8967	1.1398	**0.0462**	**0.005**
Population-based MCMC	1.9859	1.1891	0.9944	0.8892

Cuadro 3.3: Parámetros estimados por los métodos Metropolis-Hastings y population-based MCMC a partir de la moda de la distribución posterior para el modelo (3.2).

Modelo (3.2)	a	b	c	d
Reales	2	1.2	1	0.9
Metropolis-Hastings	2.2097	1.2937	0.8841	0.8061
Population-based MCMC	2.2336	1.2446	0.8447	0.7874

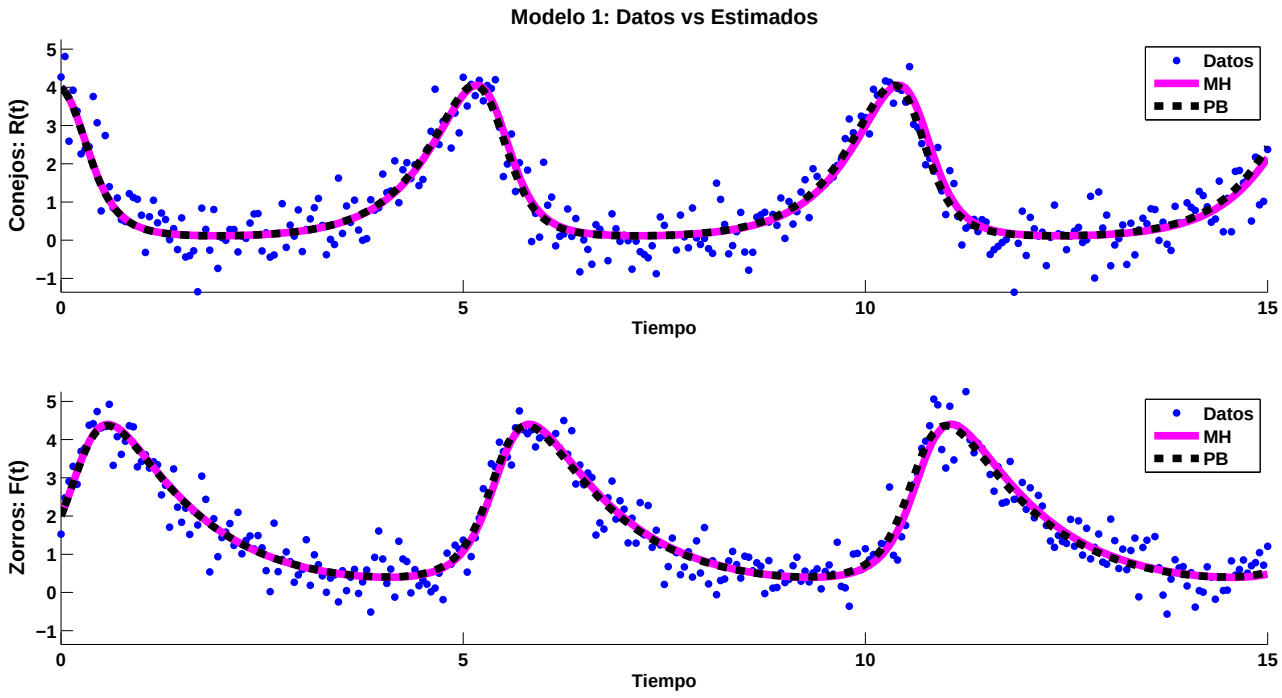

Figura 3.5: Soluciones estimadas para el modelo (3.1) en la primera réplica. Soluciones a partir de los métodos Metropolis-Hastings (MH, línea magenta continua) y population-based MCMC (PB, línea negra cortada).

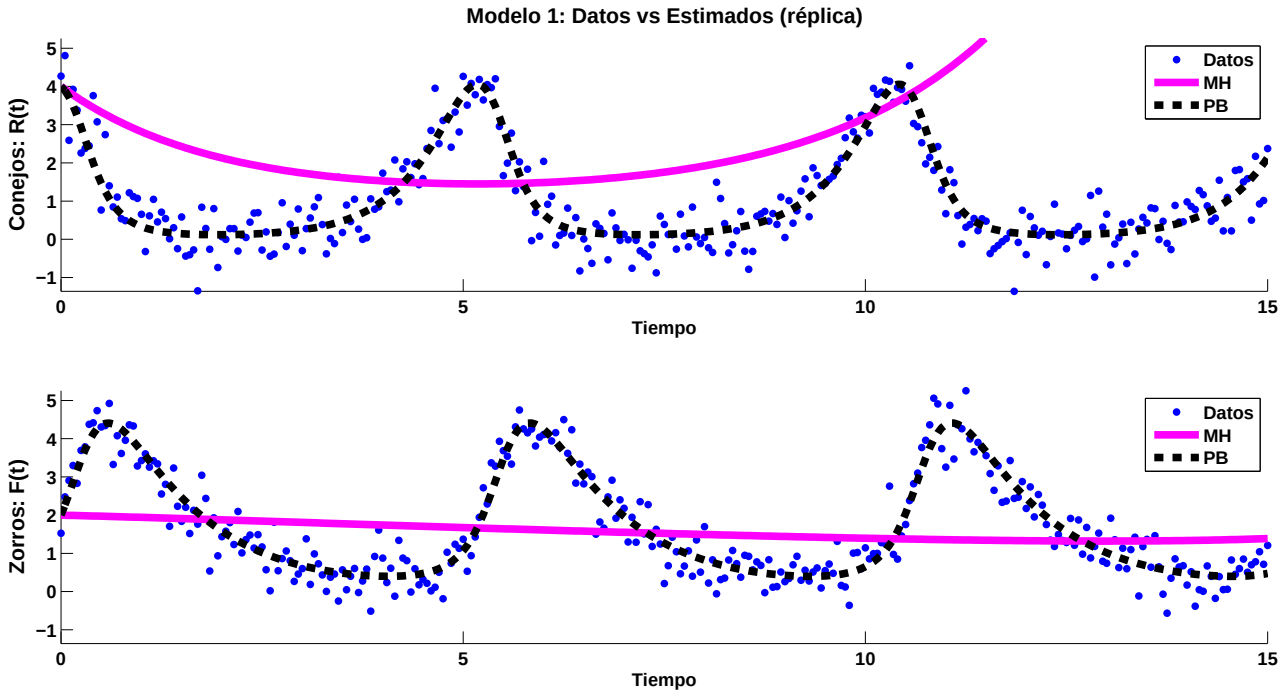

Figura 3.6: Soluciones estimadas para el modelo (3.1) en la segunda réplica. Soluciones a partir de los métodos Metropolis-Hastings (MH, línea magenta continua) y population-based MCMC (PB, línea negra cortada).

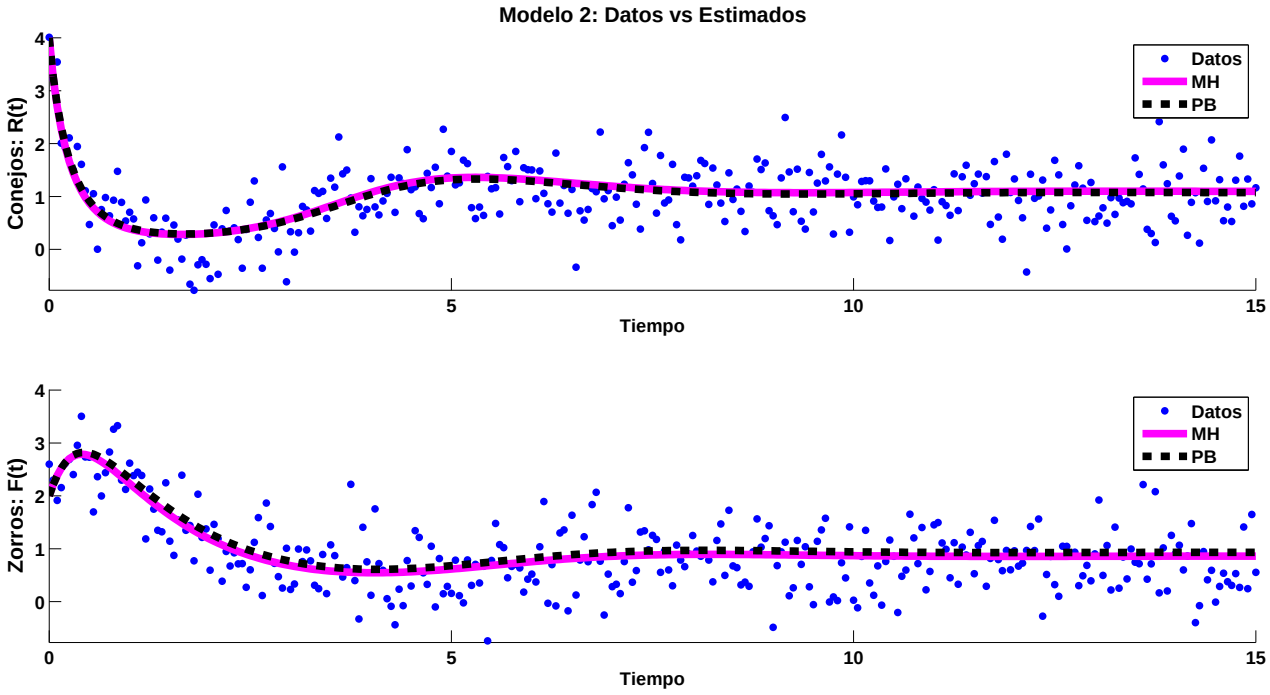

Figura 3.7: Soluciones estimadas para el modelo (3.2) en la primera réplica. Soluciones a partir de los métodos Metropolis-Hastings (MH, línea magenta continua) y population-based MCMC (PB, línea negra cortada).

3.4. Resultados

El método Metropolis-Hastings es sensible al enfrentarse a modelos complejos, en este caso ante una dinámica oscilatoria como lo es el modelo (3.1). El inconveniente del local-trap que presenta este método genera problemas graves de estimación en las soluciones como lo muestra la Figura 3.6, donde al arrojar valores de parámetros (parámetros c y d) producto de un estancamiento en regiones subóptimas (producto de la multimodalidad), genera soluciones completamente alejadas del comportamiento de los datos experimentales. El método Metropolis-Hastings resulta ser impredecible ante modelos complejos, pues si bien podría alcanzar la moda óptima en una estimación, es capaz, de en una estimación distinta, saltar a una moda subóptima.

Contrastando en la estimación se encuentra el método population-based MCMC, que si bien genera distribuciones multimodales ante modelos complejos (bimodales en el caso del modelo (3.1)), distingue perfectamente entre la moda óptima y subóptima independientemente de cuantas veces se realice la estimación; situación que no ocurre, como ya se observó, con el método Metropolis-Hastings. Como resultado, el método population-based MCMC logra descartar modas subóptimas logrando una mejor aproximación al valor real, independientemente de la complejidad del modelo.

En síntesis, el método Metropolis-Hastings resulta ser un algoritmo eficiente ante modelos de baja complejidad y donde la información a priori es suficiente; mientras que para modelos complejos y cuando se posee poca información, el método population-based MCMC resulta tener un mejor desempeño en la estimación de parámetros corroborando lo mencionado en general por Vyshemirsky y Girolami (2008).

4. Efecto del Ruido en Datos y de la Complejidad de Sistemas Dinámicos Sobre la Estimación Bayesiana

Introducción

En la sección anterior se comparó un par de métodos de estimación Bayesianos y se concluyó que el algoritmo population-based MCMC se desempeña de mejor manera ante sistemas dinámicos complejos y con falta de información del fenómeno. Tras haber optado por un método de estimación, lo que sigue es determinar qué efectos tiene el ruido muestral sobre la estimación de parámetros, así como el efecto de la complejidad del modelo propuesto. Para tal fin, la idea es considerar distintos niveles de ruido además de un modelo dinámico con la cualidad de que su comportamiento sea altamente complejo, y observar el efecto de tales circunstancias sobre las distribuciones posteriores de los parámetros. Específicamente, se utilizará el sistema de Lorenz como modelo representativo de este análisis.

Uno de los sistemas de ecuaciones diferenciales más famosos en el campo de la meteorología es el denominado sistema de Lorenz, formulado en 1963 debe su nombre a E. N. Lorenz que buscaba construir un modelo que explicara la convección atmosférica (Hirsch *et al*., 2004 y Heinz-Otto *et al*., 2004). Debido a su naturaleza, el sistema de Lorenz constituye un tipo especial de modelo cuyo comportamiento resulta ser caótico en su dinámica. La definición de caos en sistemas determinísticos descrita por Heinz-Otto *et al*. (2004) se refiere a que un mismo sistema genera un comportamiento "*aleatorio*" como parte fundamental de su dinámica y que no importando la cantidad de información suministrada al modelo, tal comportamiento no desaparece. Esta *"aleatoriedad fundamental"* es a lo que se denomina caos. Sin embargo, el uso de la palabra aleatorio únicamente cobra sentido al representar el comportamiento impredecible del fenómeno, pues el caos es determinístico, es decir, está regido por reglas fijas que no envuelven ningún elemento aleatorio. A pesar de que el sistema de Lorenz es un intento de modelar una dinámica tan complicada como lo es el clima, el sistema resultante utiliza únicamente ecuaciones diferenciales ordinarias de primer orden. La idea sobre la que se basa el sistema es simple, a partir de un único fluido, éste es calentado desde abajo y por lo tanto asciende, posteriormente es enfriado desde arriba y por lo tanto desciende (Hirsch *et al*., 2004).

El sistema de Lorenz consta de tres ecuaciones ordinarias de primer orden y tres parámetros ($\theta_0 = (\mathbf{s}, \mathbf{b}, \mathbf{r})$) como se muestra a continuación (Scheinerman, 1996):

$$\begin{aligned} \frac{dX}{dt} &= \mathbf{s} \cdot (Y - X) \\ \frac{dY}{dt} &= \mathbf{r} \cdot X - Y - XZ \\ \frac{dZ}{dt} &= XY - \mathbf{b} \cdot Z \end{aligned}$$

A partir de los valores de los parámetros **s**, **b** y **r**, se consideraron dos esquemas del sistema anterior, un estado caótico y un estado no caótico. Para el estado no caótico pero cíclico convergente los parámetros fueron:

$$\theta_0 = (\mathbf{s} = 10, \mathbf{b} = 2.6666, \mathbf{r} = 18.3) \tag{4.1}$$

Tal configuración corresponde a un estado de convección estacionario, y la solución al sistema de Lorenz correspondiente se muestra en la Figura 4.1, donde se observa un claro comportamiento oscilatorio periódico convergente. Por otro lado, para el estado caótico, los parámetros correspondientes son:

$$\theta_0 = (\mathbf{s} = 10, \mathbf{b} = 2.6666, \mathbf{r} = 28) \tag{4.2}$$

Estos son los parámetros originales que llevaron al descubrimiento por parte de Lorenz de las implicaciones del caos, tal configuración determina una dinámica oscilatoria pero irregular no convergente en cada estado como se muestra en la Figura 4.2 (Hirsch *et al.*, 2004).

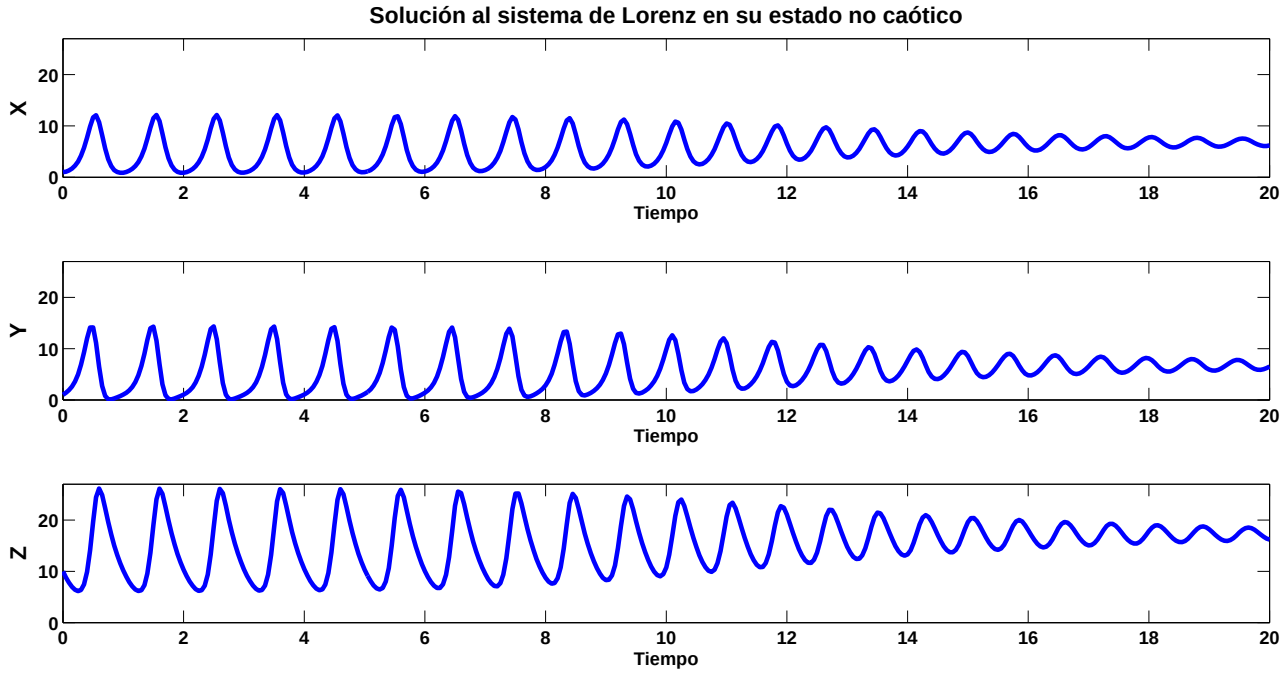

Figura 4.1: Soluciones para el sistema de Lorenz en su estado no caótico correspondiente a los parámetros (4.1).

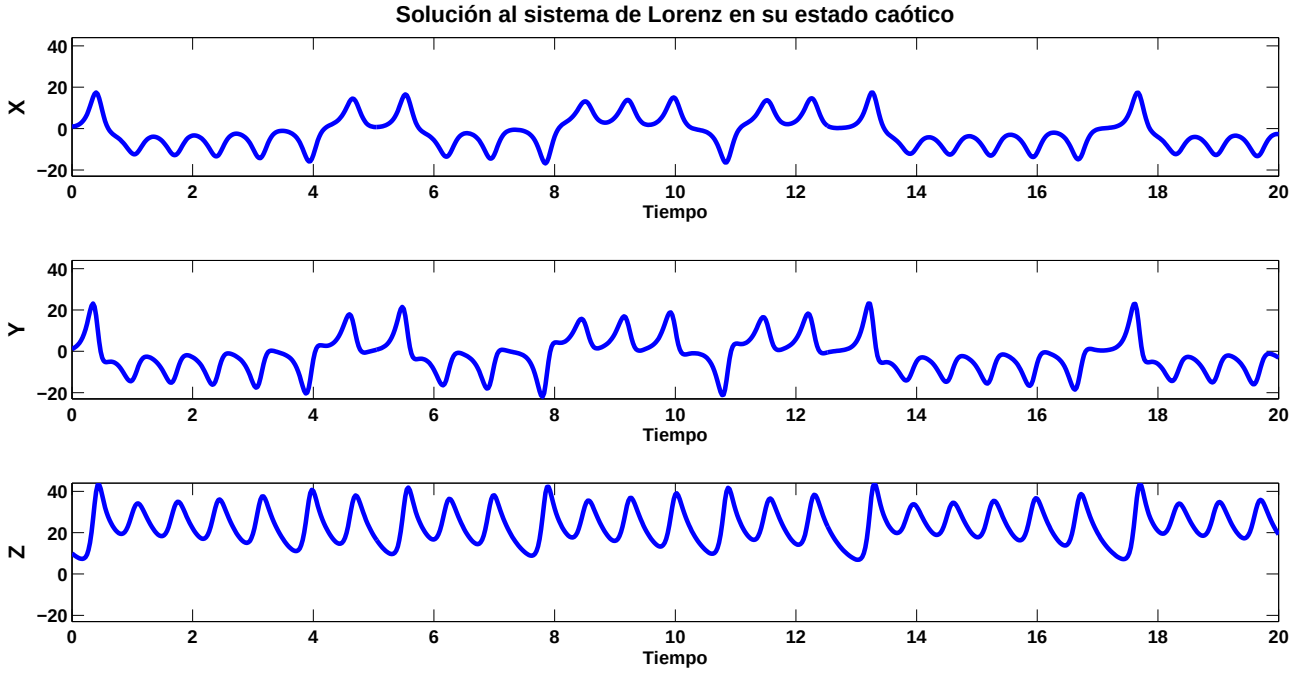

Figura 4.2: Soluciones para el sistema de Lorenz en su estado caótico correspondiente a los parámetros (4.2).

4.1. Simulación y estimación

Para cada conjunto de parámetros se calcularon las soluciones del sistema de Lorenz con condiciones iniciales dadas por (4.3), en un intervalo de tiempo $t = [0, 20]$.

$$\begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 10 \end{bmatrix} \tag{4.3}$$

Al conjunto de datos así construido se le adicionó ruido gaussiano con media cero y desviación estándar σ de 0, 0.5, 1, 2 y 4, respectivamente, a la que referiremos en adelante como nivel de ruido. Para el caso de $\sigma = 0$, se están considerando los datos sin ruido añadido, obtenidos directamente de la simulación bajo los parámetros dados por (4.1) y (4.2), Figuras 4.1 y 4.2, respectivamente.

La condición inicial de cada variable se conservó fija, es decir, no se le añadió ruido alguno. Para la estimación se utilizó el método population-based MCMC por medio del software BioBayes. En ambos casos (estado caótico y no caótico) se dieron distribuciones a priori informativas de tipo Gaussianas con media igual al valor real del parámetro y con una desviación estándar σ de 0.1. Se fijó un límite de Burn-In de 200,000 iteraciones con un Thinning muestral de 5 y una muestra posterior de tamaño 2000. Para la estimación en el sistema con estado no caótico se utilizaron 8 cadenas paralelas e igual número de muestreadores independientes. En el caso caótico se utilizaron 10 cadenas en paralelo y 5 muestreadores independientes.

4.2. Sistema de Lorenz: estado no caótico

El Cuadro 4.1 muestra los parámetros estimados por la moda de las distribuciones posteriores de los parámetros del sistema de Lorenz en su estado no caótico a distintos niveles de ruido. Este estimador dado por la moda no presenta problema de estimación. Las Figuras 4.3, 4.4 y 4.5 muestran las distribuciones posteriores de los parámetros **s**, **b** y **r**; a distintos niveles de ruido muestral para el sistema de Lorenz en su estado no caótico.

Cuadro 4.1: Parámetros reales y estimados a partir de datos a distintos niveles de ruido con desviación estándar sigma para el sistema de Lorenz en su estado no caótico.

Parámetro	Real	Sin ruido	Ruido 0.5	Ruido 1	Ruido 2	Ruido 4
s	10	9.9901	10.01	10.1498	10.0334	10.0253
b	2.6666	2.6922	2.6668	2.6549	2.667	2.6704
r	18.3	18.2432	18.303	18.2792	18.3088	18.3055

4.2.1. Estimación

Parámetro s

La Figura 4.3 muestra las distribuciones posteriores (misma escala en el eje horizontal para los distintos niveles de ruido), del parámetro **s** a distintos niveles de ruido. En la Figura 4.3 para los datos sin ruido añadido ($\sigma = 0$, primer fila) y datos con ruido $\sigma = 2$ y $\sigma = 4$ (filas 4 y 5), se observa que las distribuciones posteriores presentan un comportamiento claramente unimodal y poseen intervalos de probabilidad (al 85 %) reducidos alrededor de la moda (línea roja cortada). Una situación contraria se presenta para las distribuciones con ruido σ de 0.5 y 1 (filas 2 y 3), que poseen intervalos de probabilidad muy extensos, además de que en el caso de $\sigma = 1$ (fila 3) tal intervalo no contiene al valor real del parámetro (línea negra punteada), aún cuando el intervalo es más estrecho que en el caso de $\sigma = 0.5$. Cabe señalar que el caso de $\sigma = 0$ posee los intervalos de probabilidad más extensos.

Parámetro b

La Figura 4.4 muestra las distribuciones posteriores (misma escala en el eje horizontal para los distintos niveles de ruido) del parámetro **b**, a distintos niveles de ruido. En general se observa que independientemente del nivel de ruido, los intervalos de probabilidad al 85 % son bastante reducidos alrededor de la moda de la distribución. Sin embargo, observando más de cerca se tiene que para la distribución posterior correspondiente a σ de 0.5 y 1 (filas 2 y 3), presentan zonas de alta densidad en sus colas izquierdas. En general salvo lo anterior mencionado, se observan distribuciones unimodales y la moda se encuentra cerca del valor real del parámetro, excepto para los casos de σ igual a 0 y 1, que presentan sus estimadores (líneas rojas cortadas) relativamente alejados del valor real (línea negra punteada). Nuevamente para el caso de $\sigma = 0$, los intervalos de probabilidad son más extensos.

Parámetro r

La Figura 4.5 muestra las distribuciones posteriores escaladas (misma escala en el eje horizontal para los distintos niveles de ruido) del parámetro **r**, a distintos niveles de ruido. Como en el caso del parámetro **s**, las distribuciones correspondientes a los datos con ruido σ de 0.5 y 1 (filas 2 y 3) poseen regiones de alta densidad en las colas derechas de las distribuciones, por lo que los intervalos de probabilidad son extensos. En general se tienen distribuciones unimodales aunque se presenta una clara bimodalidad en al caso de $\sigma = 0.5$ (fila 2). Nuevamente es persistente que la distribución para el caso de $\sigma = 0$ es unimodal y de intervalos de probabilidad extensos respecto a las demás distribuciones.

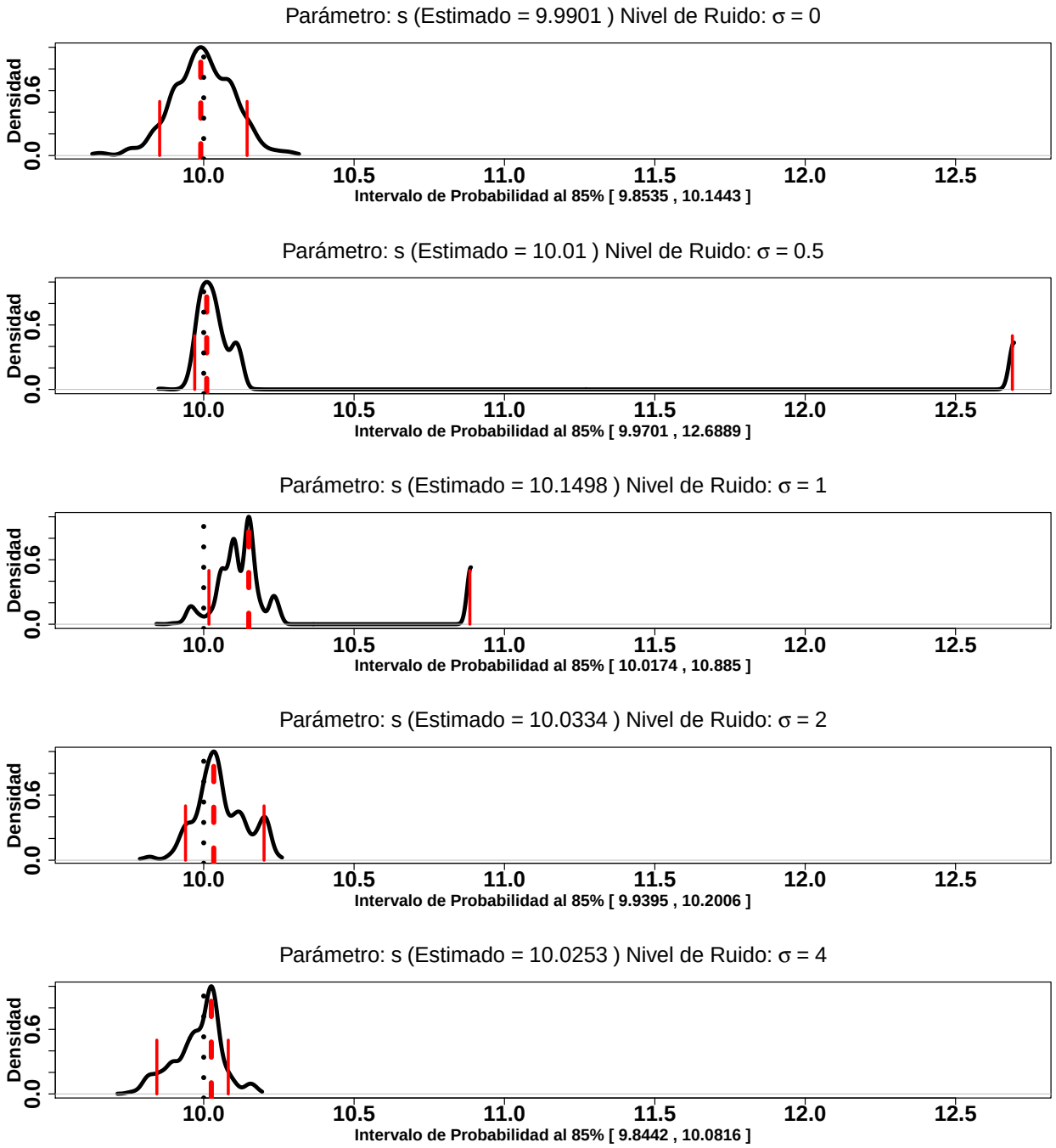

Figura 4.3: Por filas, la distribución posterior del parámetro **s** del sistema de Lorenz en su estado no caótico a distintos niveles de ruido con desviación estándar σ de 0, 0.5, 1, 2, y 4, respectivamente. Valor real (línea negra punteada), valor estimado (línea roja cortada) e intervalo de probabilidad al 85 % (líneas rojas sólidas).

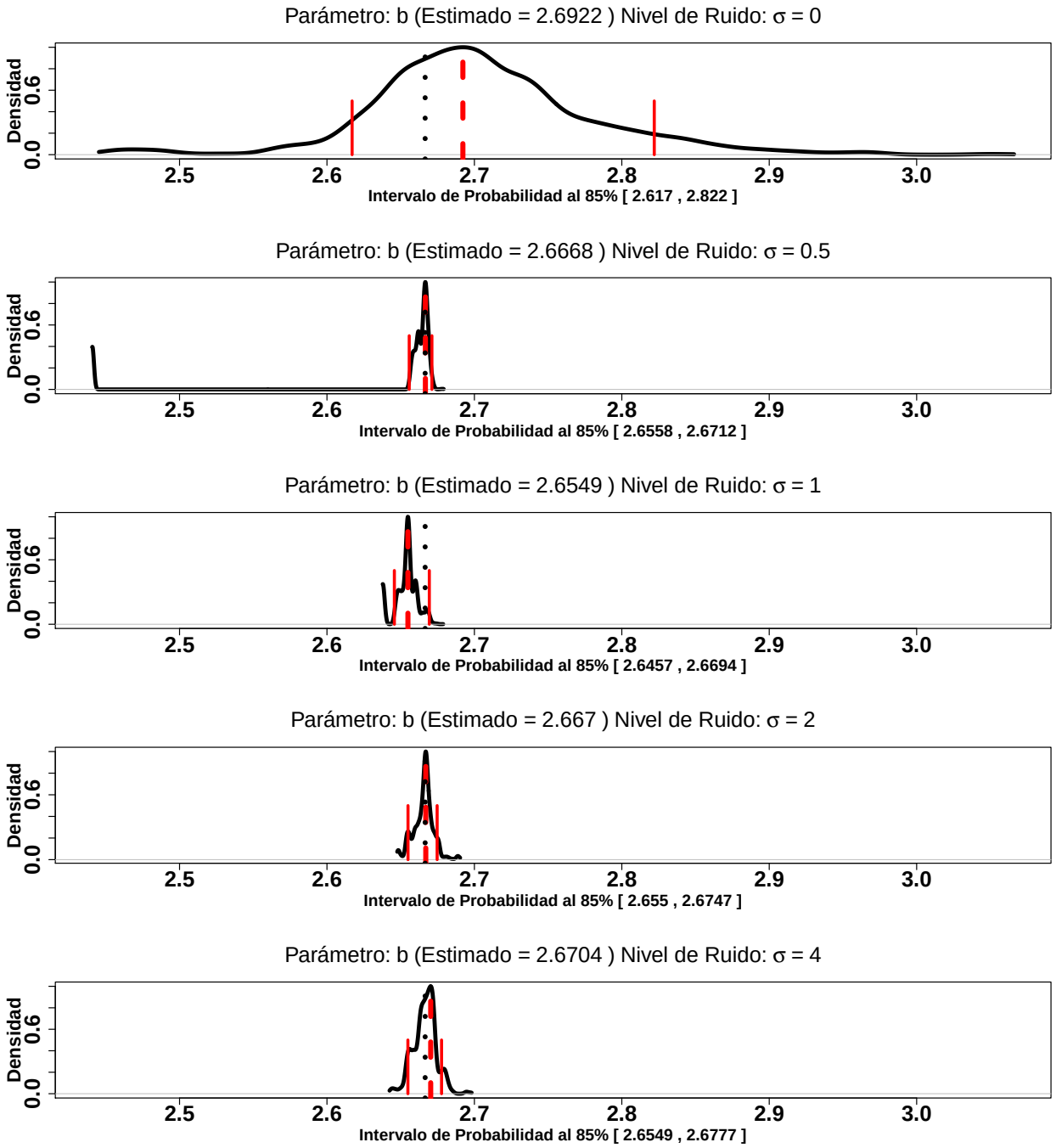

Figura 4.4: Por filas, la distribución posterior del parámetro **b** del sistema de Lorenz en su estado no caótico a distintos niveles de ruido con desviación estándar σ de 0, 0.5, 1, 2, y 4, respectivamente. Valor real (línea negra punteada), valor estimado (línea roja cortada) e intervalo de probabilidad al 85 % (líneas rojas sólidas).

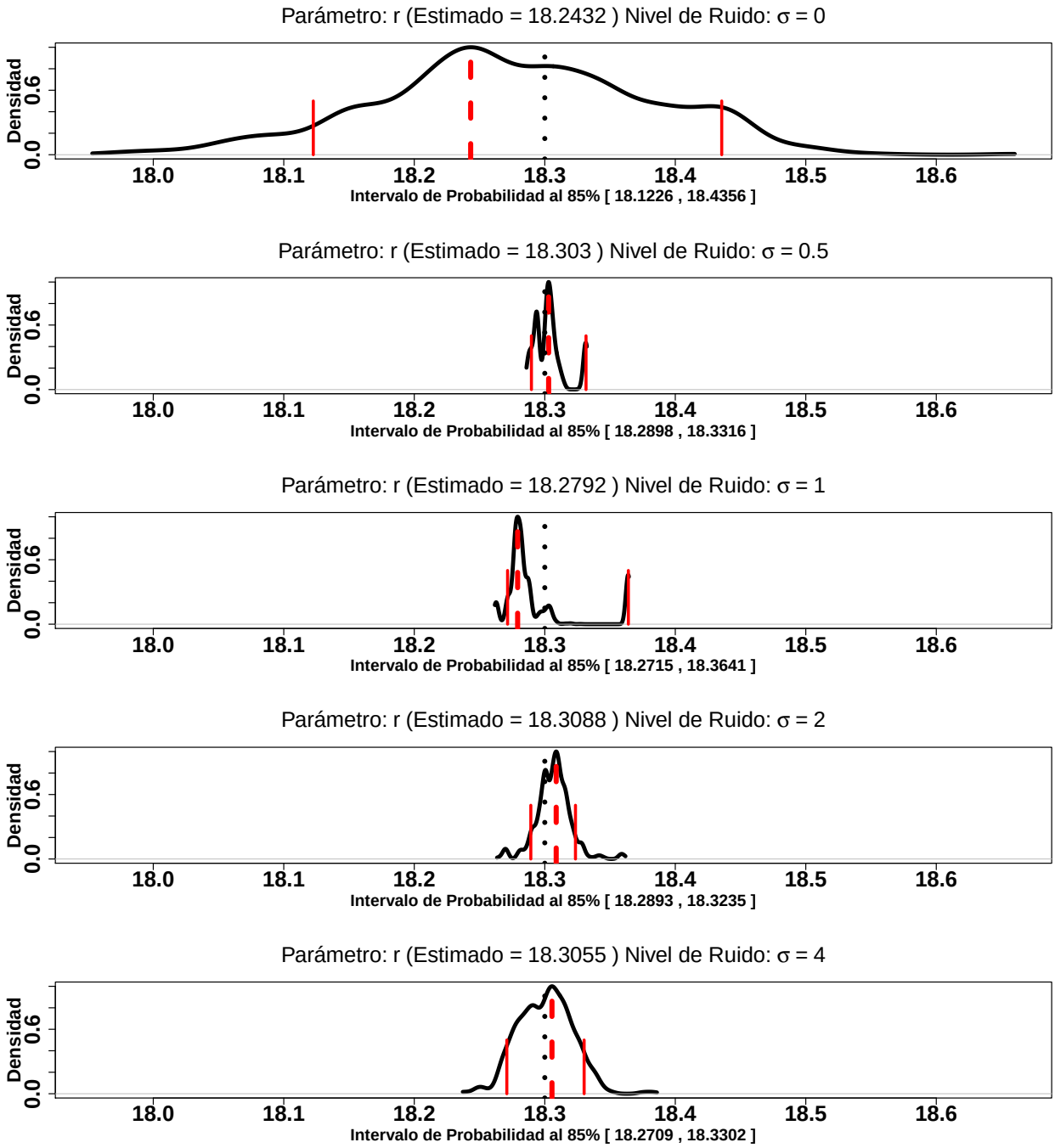

Figura 4.5: Por filas, la distribución posterior del parámetro **r** del sistema de Lorenz en su estado no caótico a distintos niveles de ruido con desviación estándar σ de 0, 0.5, 1, 2, y 4, respectivamente. Valor real (línea negra punteada), valor estimado (línea roja cortada) e intervalo de probabilidad al 85 % (líneas rojas sólidas).

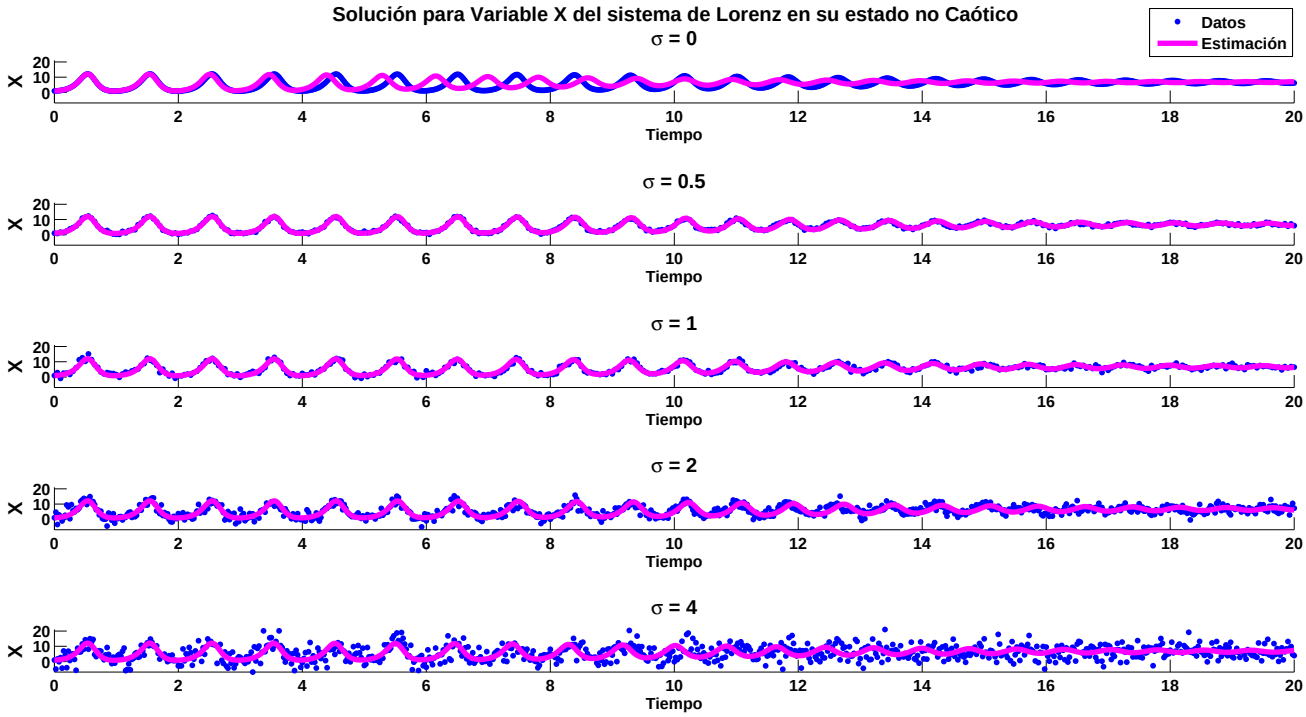

Figura 4.6: Por filas, la solución estimada (línea sólida magenta) y los datos con ruido (puntos azules) del sistema de Lorenz en su estado no caótico a distintos niveles de ruido (σ =0, 0.5, 1, 2 y 4) para la variable X.

4.2.2. Soluciones

Las Figuras 4.6, 4.7 y 4.8, corresponden a las soluciones estimadas para las variables X, Y y Z, del sistema de Lorenz en su estado no caótico a partir de los parámetros estimados por la moda de las distribuciones posteriores, Cuadro 4.1. Para las variables X, Y y Z para el nivel de ruido $\sigma = 0$ (fila 1 de las Figuras 4.6, 4.7 y 4.8) se observa un desfase temprano de la solución estimada (línea sólida magenta) respecto a los datos (puntos azules). Sin embargo la tendencia de la solución estimada concuerda con la tendencia de los datos en el largo plazo. Para el caso de las soluciones para datos con ruido de desviación estándar σ igual a 0.5, 1 y 2, las soluciones estimadas no presentan el desfase como en el caso anterior, y el ajuste así como la tendencia es similar a los datos experimentales para cada variable. Para las soluciones para los datos con ruido de desviación estándar σ de 4 (ruido alto), se observa un desfase como en el primer caso ($\sigma = 0$), pero ahora en un estado tardío de la dinámica.

Finalmente, la Figura 4.9 muestra los planos fase (gráfico de las tres variables) para los distintos niveles de ruido así como el plano fase para los parámetros reales. Se observa una similitud entre cada plano fase, aunque es evidente que para el caso de $\sigma = 0$, la dinámica se acelera alrededor del punto de equilibrio, por lo que se observa una dinámica más concentrada alrededor del atractor.

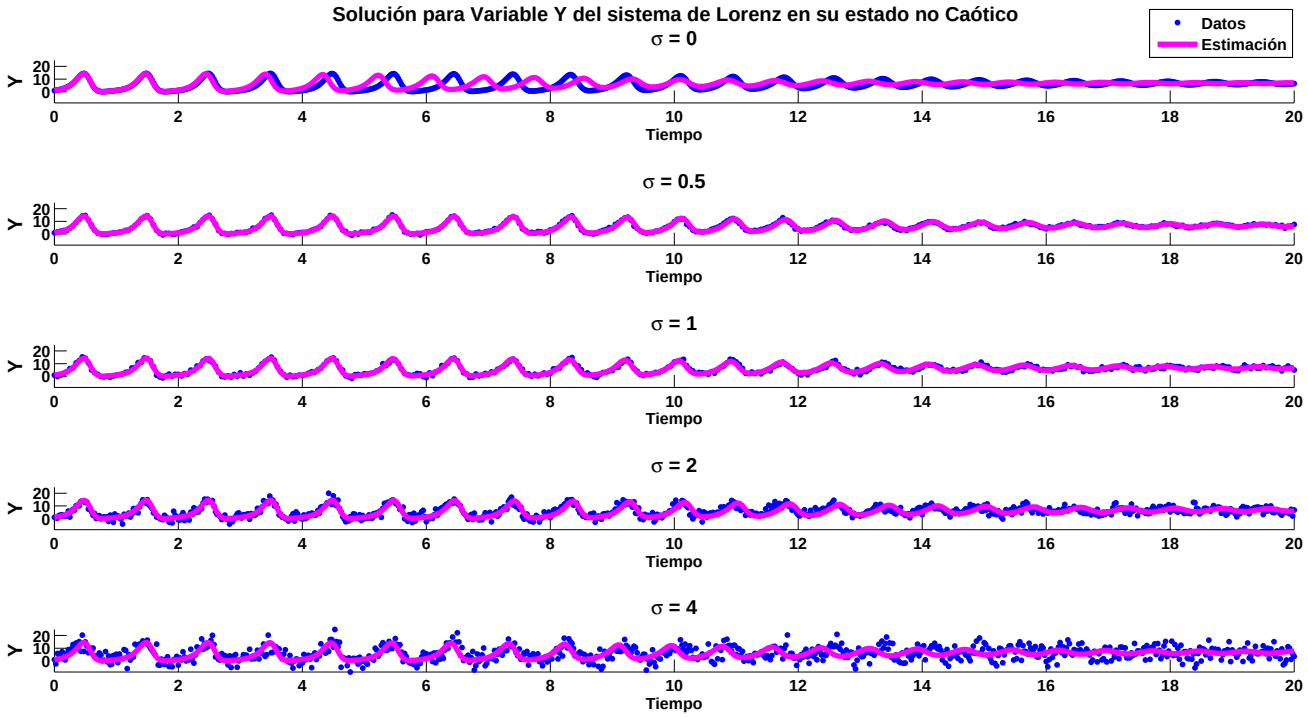

Figura 4.7: Por filas, la solución estimada (línea sólida magenta) y los datos con ruido (puntos azules) del sistema de Lorenz en su estado no caótico a distintos niveles de ruido (σ =0, 0.5, 1, 2 y 4) para la variable Y.

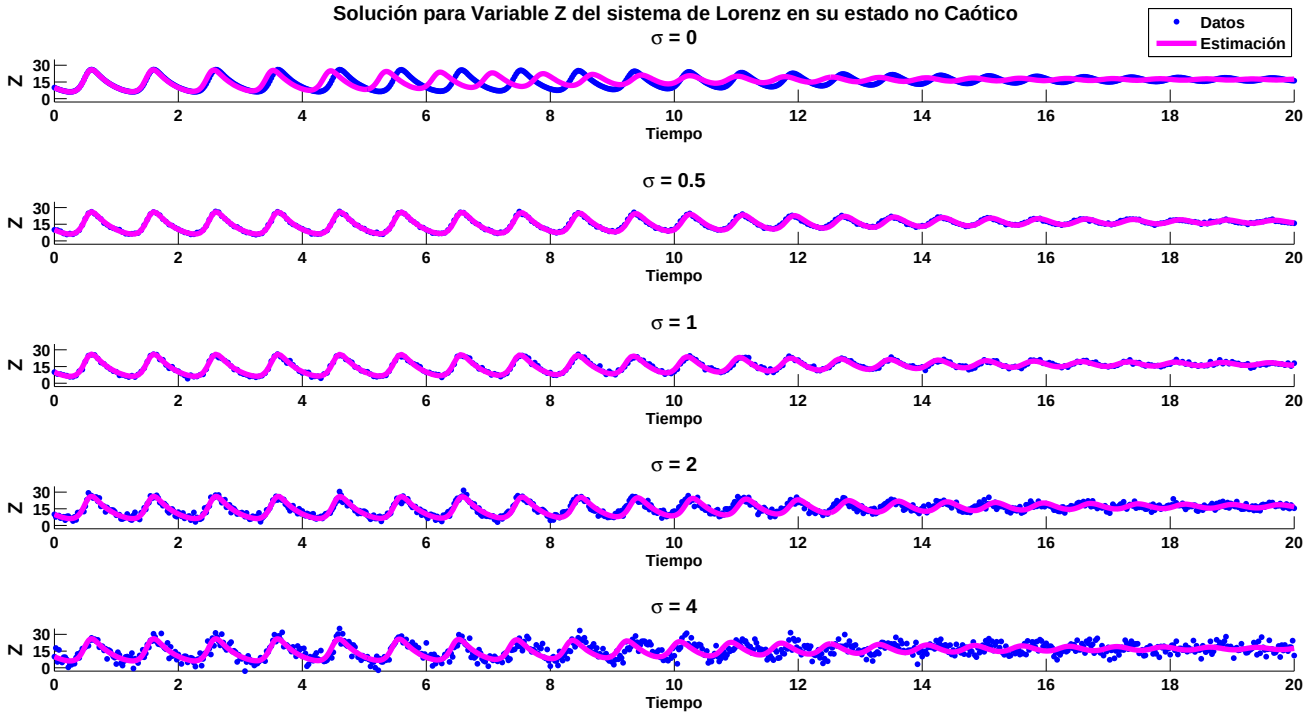

Figura 4.8: Por filas, la solución estimada (línea sólida magenta) y los datos con ruido (puntos azules) del sistema de Lorenz en su estado no caótico a distintos niveles de ruido (σ =0, 0.5, 1, 2 y 4) para la variable Z.

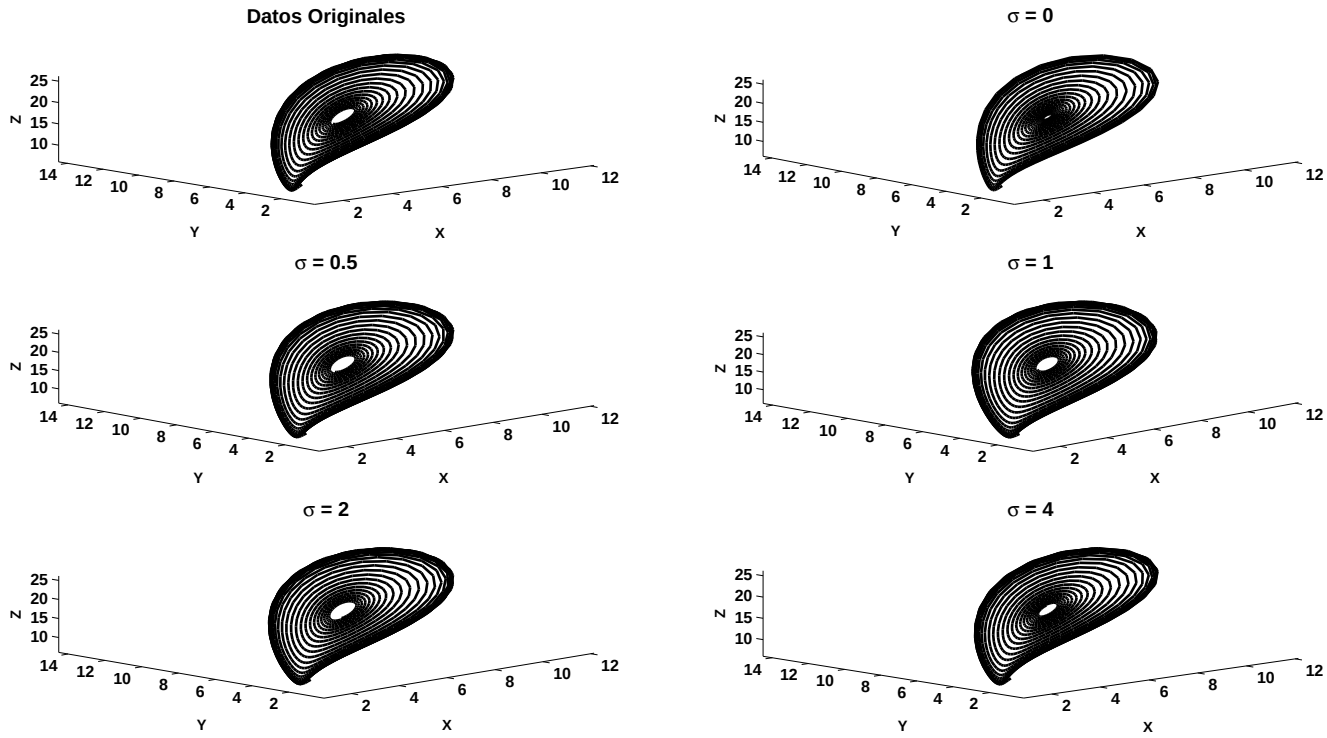

Figura 4.9: Planos fase de los datos originales (primer figura esquina superior izquierda) y de la solución estimada (figuras restantes) del sistema de Lorenz (estado no caótico) a distintos niveles de ruido.

Cuadro 4.2: Parámetros reales y estimados a partir de datos a distintos niveles de ruido con desviación estándar sigma para el sistema de Lorenz en su estado caótico.

Parámetro	Real	Sin ruido	Ruido 0.5	Ruido 1	Ruido 2	Ruido 4
s	10	10.0281	10.0704	9.9029	10.0218	10.0052
b	2.6666	2.669	0.6979239	0.3659147	2.6644	2.6669
r	28	27.8919	27.9297	28.1687	27.9838	28.0468

4.3. Sistema de Lorenz: Estado caótico

El Cuadro 4.2 muestra los parámetros estimados por la moda de las distribuciones posteriores de los parámetros del sistema de Lorenz en su estado caótico a distintos niveles de ruido. Las Figuras 4.10, 4.12 y 4.14 muestran las distribuciones posteriores de los parámetros **s**, **b** y **r** a distintos niveles de ruido muestral para el sistema de Lorenz en su estado caótico.

4.3.1. Estimación

Parámetro s

Las Figuras 4.10 y 4.11 muestran las distribuciones posteriores escaladas (misma escala en el eje horizontal para los distintos niveles de ruido) y no escaldas (cada distribución con

su propia escala), respectivamente, del parámetro **s**, a distintos niveles de ruido. En general se observan distribuciones multimodales con regiones de alta densidad en las colas de las distribuciones, Figura 4.11, sin embargo el estimador (línea roja cortada) se encuentra cercano al valor real del parámetro (línea negra punteada). Tal multimodalidad se hace más evidente en el nivel de ruido $\sigma = 0.5$. En la Figura 4.10 se observa que las distribuciones posteriores son más concentradas al irse incrementando el ruido muestral.

Parámetro b

Las Figuras 4.12 y 4.13 muestran las distribuciones posteriores escaladas (misma escala en el eje horizontal para los distintos niveles de ruido) y no escaldas (cada distribución con su propia escala), respectivamente, del parámetro **b**, a distintos niveles de ruido. En la Figura 4.12 (distribuciones escaladas) se observa que para ruidos elevados (ruidos con desviación estándar σ de 2 y 4, filas 4 y 5) y para datos sin ruido (primera fila, $\sigma = 0$), las distribuciones posteriores están concentradas alrededor de la moda, que si bien existe presencia de multimodalidades (Figura 4.13) la estimación se encuentra cercana al valor real del parámetro. Algo contrario ocurre para las distribuciones correspondientes a datos con ruido de distribución estándar σ de 0.5 y 1 (Figuras 4.12 y 4.13, filas 2 y 3) se observan claramente una bimodalidad y la estimación correspondiente (línea roja cortada) se encuentra alejada de la moda alrededor del valor real del parámetro para ambos niveles de ruido (línea negra punteada).

Parámetro r

Las Figuras 4.14 y 4.15 muestran las distribuciones posteriores escaladas (misma escala en el eje horizontal para los distintos niveles de ruido) y no escaldas (cada distribución con su propia escala), respectivamente, del parámetro **r**, a distintos niveles de ruido. Al igual que el caso anterior (parámetro **b**), se puede apreciar en la Figura 4.14 (distribuciones escaladas) que para ruidos elevados (ruidos con desviación estándar σ de 2 y 4, filas 4 y 5) y para datos sin ruido (primera fila, $\sigma = 0$), las distribuciones están concentradas alrededor del valor real del parámetro (línea negra punteada), pero la multimodalidad es más evidente (Figura 4.15). Tal concentración hace que el parámetro estimado (línea roja cortada) esté cercano del valor real del parámetro. Finalmente, para las distribuciones correspondientes a datos con ruido de distribución estándar σ de 0.5 y 1 (Figuras 4.14 y 4.15, filas 2 y 3) se observa nuevamente la multimodalidad pero ahora con distribuciones más dispersas, pero la estimación se encuentra cercana al valor real del parámetro.

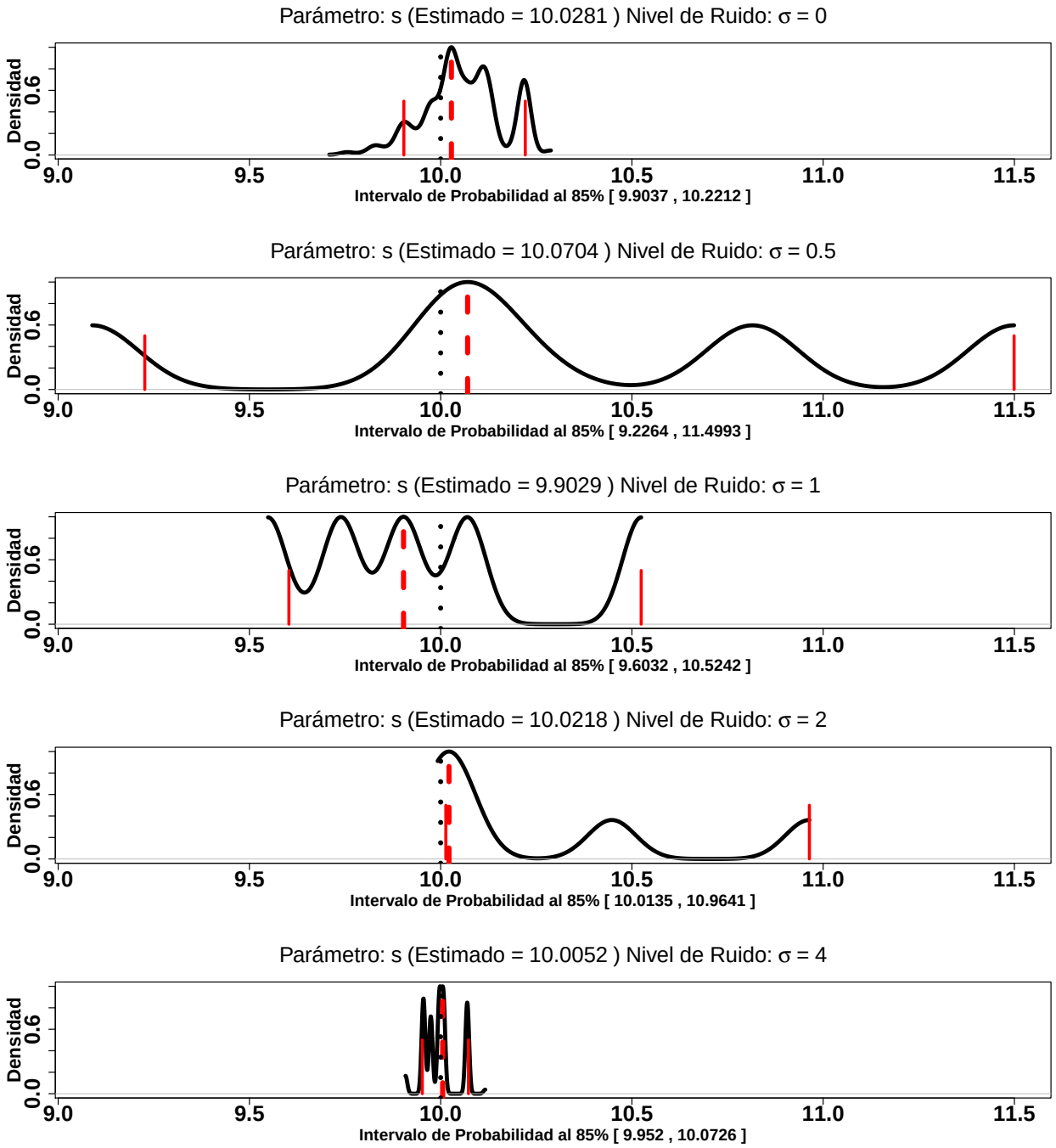

Figura 4.10: Por filas, la distribución posterior del parámetro **s** el sistema de Lorenz en su estado caótico a distintos niveles de ruido con desviación estándar σ de 0, 0.5, 1, 2, y 4, respectivamente. Valor real (línea negra punteada), valor estimado (línea roja cortada) e intervalo de probabilidad al 85 % (líneas rojas sólidas). Escalada.

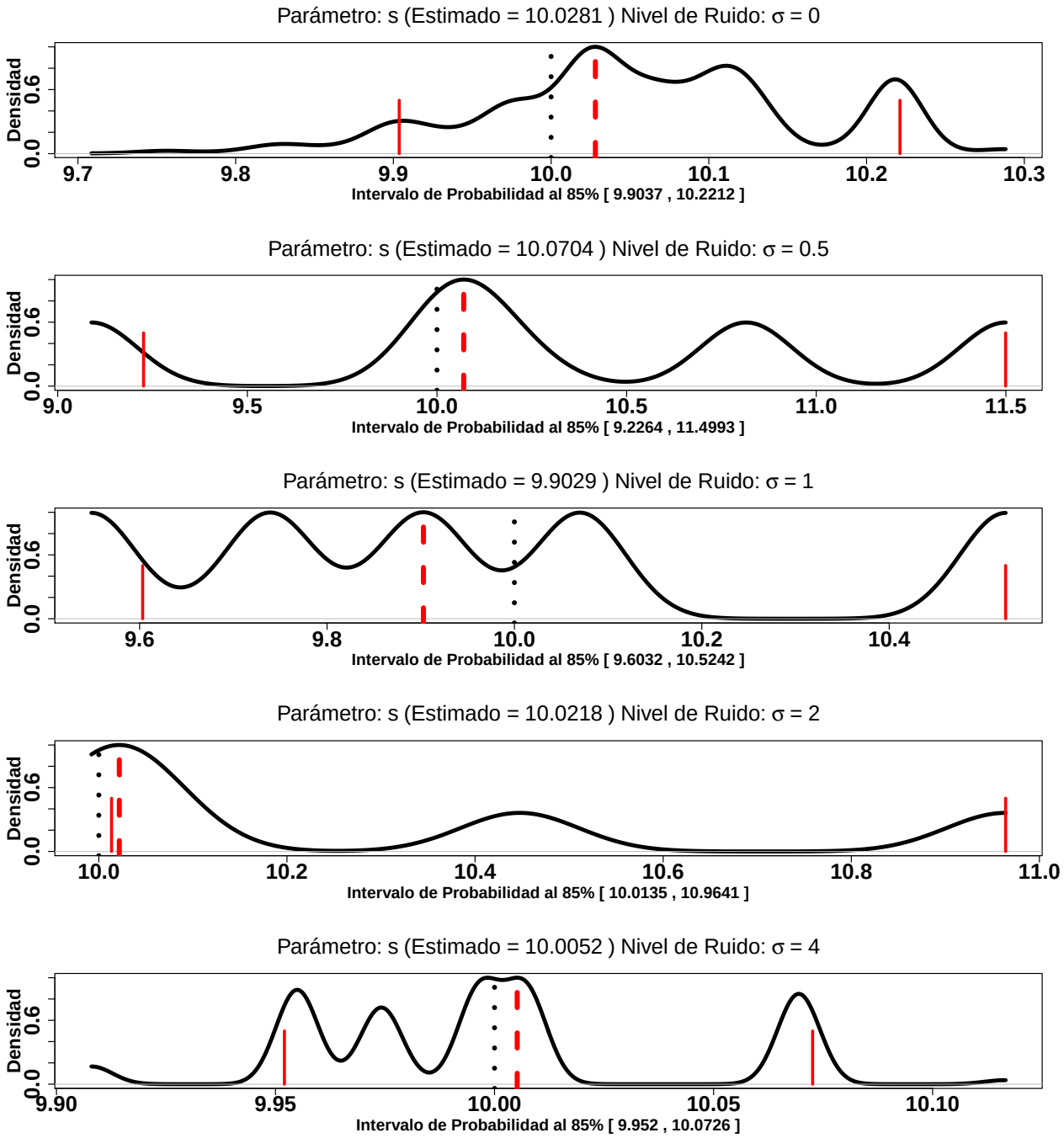

Figura 4.11: Por filas, la distribución posterior del parámetro **s** el sistema de Lorenz en su estado caótico a distintos niveles de ruido con desviación estándar σ de 0, 0.5, 1, 2, y 4, respectivamente. Valor real (línea negra punteada), valor estimado (línea roja cortada) e intervalo de probabilidad al 85 % (líneas rojas sólidas). No Escalada.

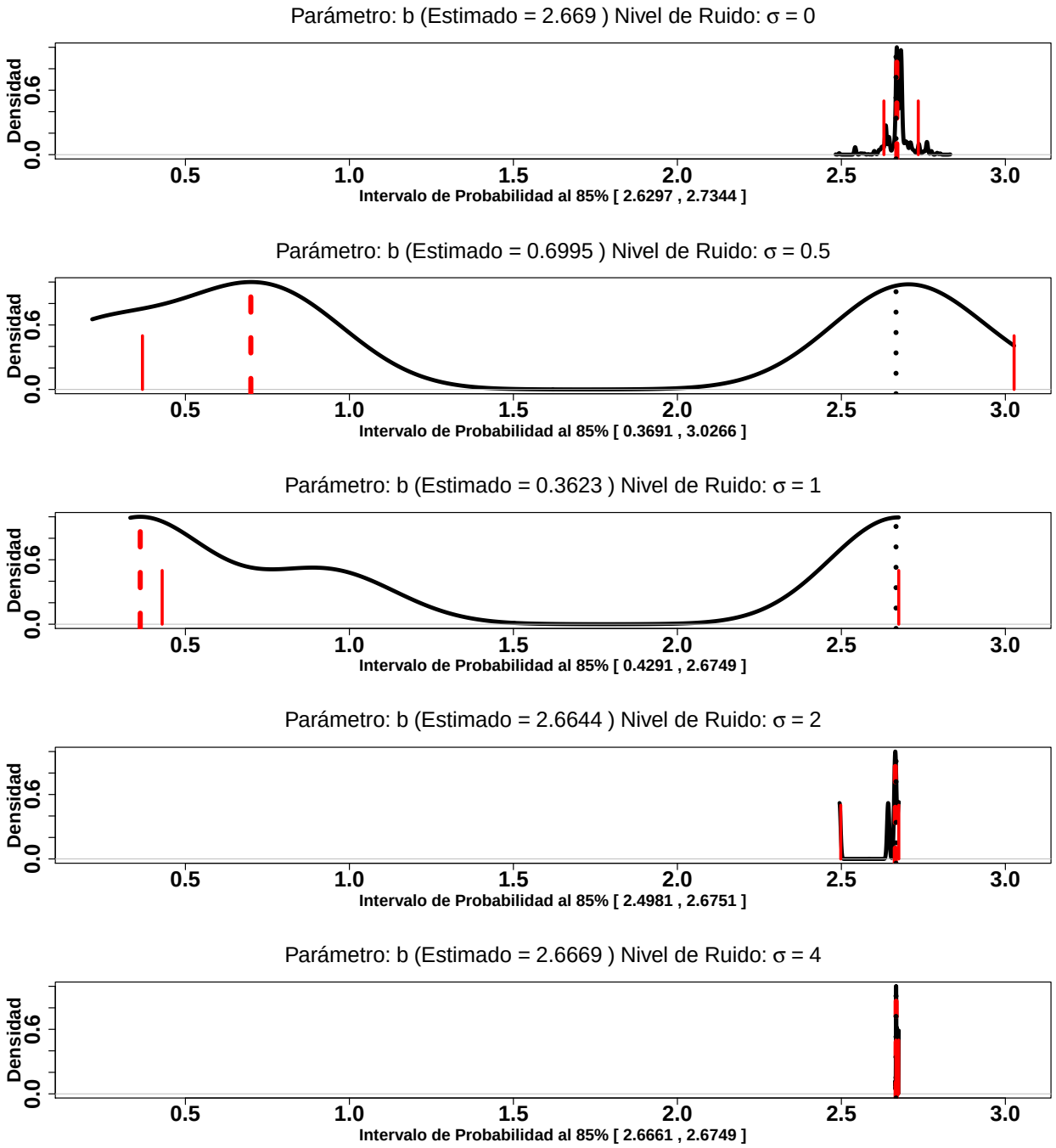

Figura 4.12: Por filas, la distribución posterior del parámetro **b** el sistema de Lorenz en su estado caótico a distintos niveles de ruido con desviación estándar σ de 0, 0.5, 1, 2, y 4, respectivamente. Valor real (línea negra punteada), valor estimado (línea roja cortada) e intervalo de probabilidad al 85 % (líneas rojas sólidas). Escalada.

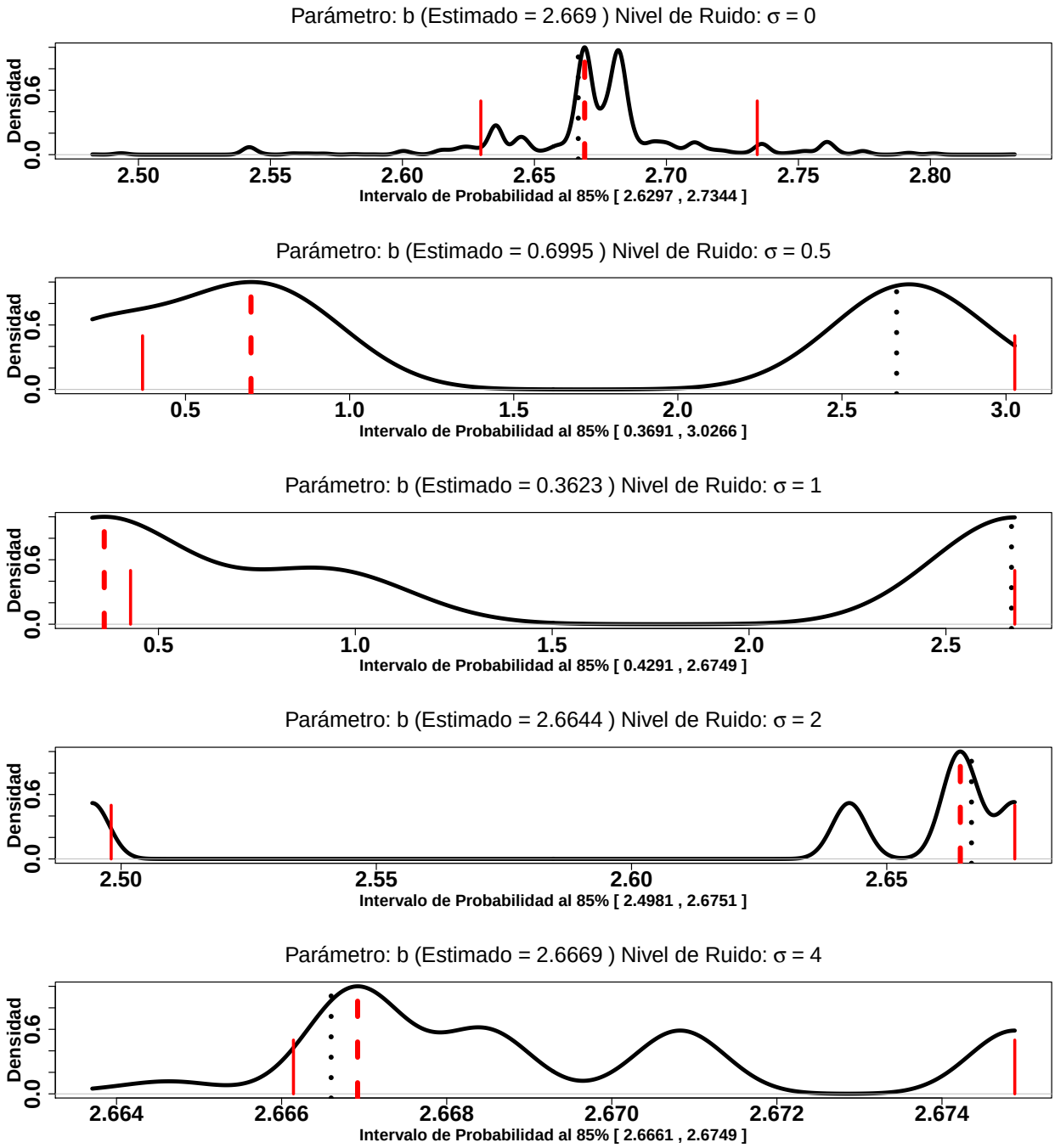

Figura 4.13: Por filas, la distribución posterior del parámetro **b** el sistema de Lorenz en su estado caótico a distintos niveles de ruido con desviación estándar σ de 0, 0.5, 1, 2, y 4, respectivamente. Valor real (línea negra punteada), valor estimado (línea roja cortada) e intervalo de probabilidad al 85 % (líneas rojas sólidas). No Escalada.

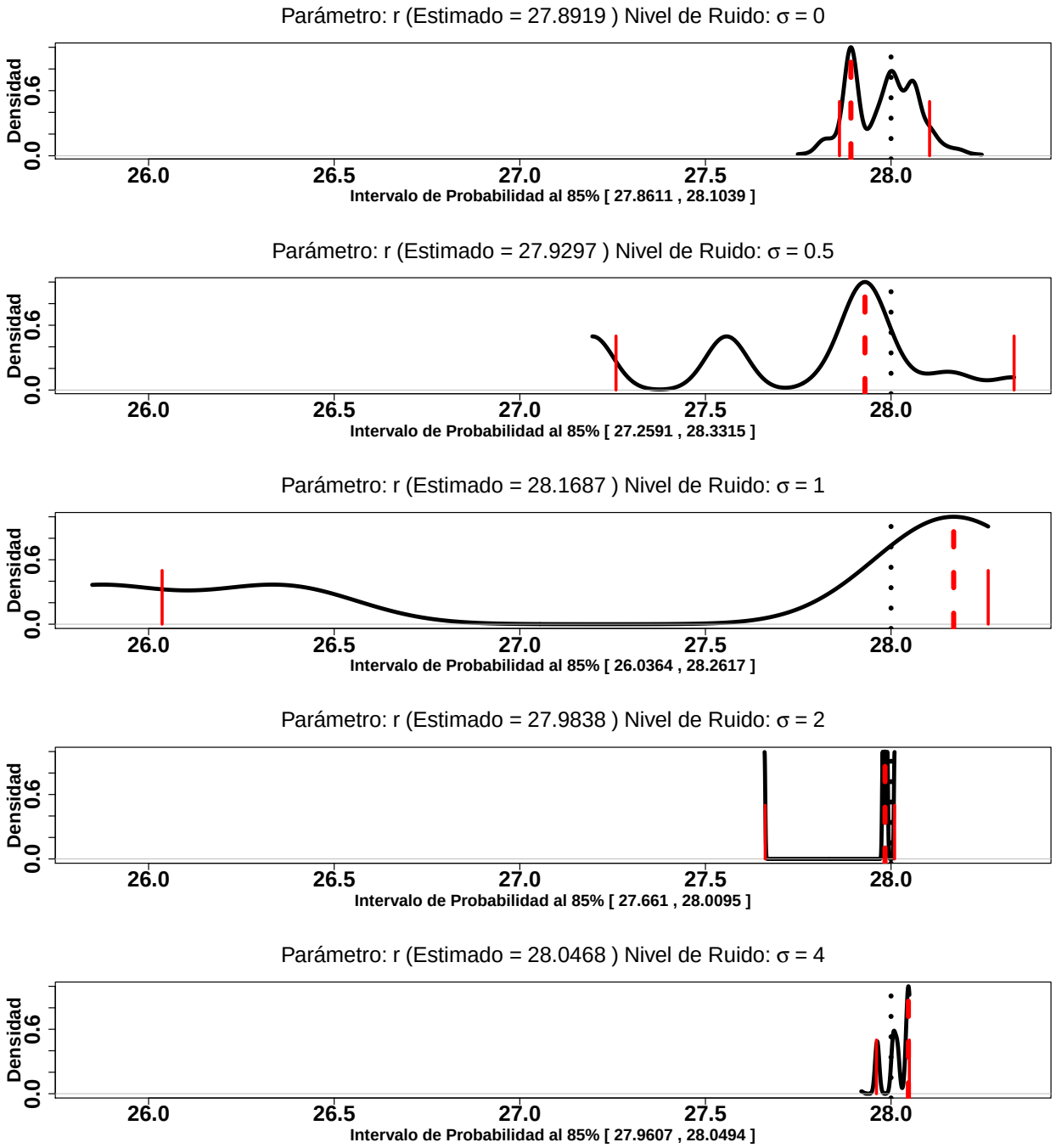

Figura 4.14: Por filas, la distribución posterior del parámetro **r** el sistema de Lorenz en su estado caótico a distintos niveles de ruido con desviación estándar σ de 0, 0.5, 1, 2, y 4, respectivamente. Valor real (línea negra punteada), valor estimado (línea roja cortada) e intervalo de probabilidad al 85 % (líneas rojas sólidas). Escalada.

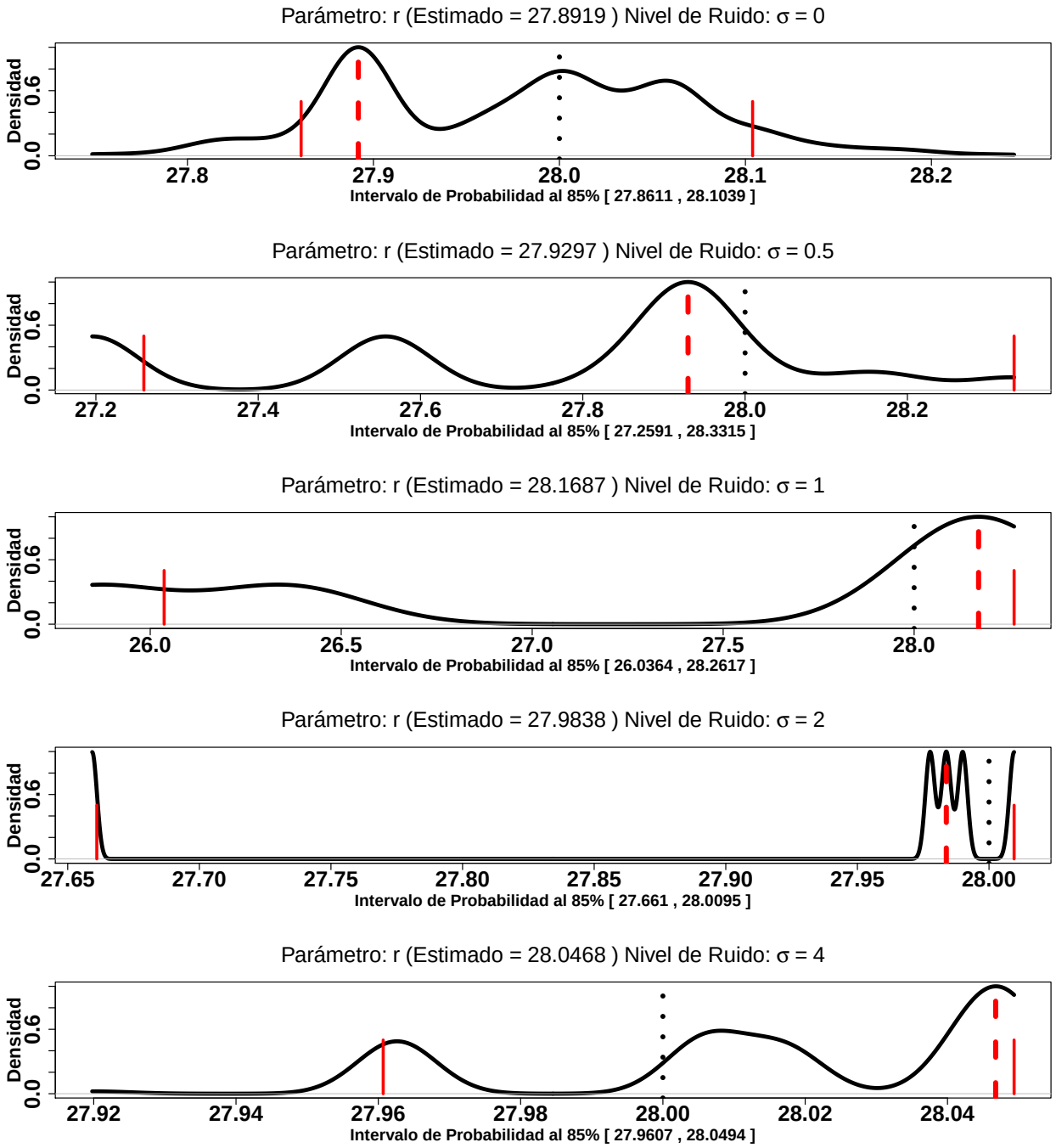

Figura 4.15: Por filas, la distribución posterior del parámetro **r** el sistema de Lorenz en su estado caótico a distintos niveles de ruido con desviación estándar σ de 0, 0.5, 1, 2, y 4, respectivamente. Valor real (línea negra punteada), valor estimado (línea roja cortada) e intervalo de probabilidad al 85 % (líneas rojas sólidas). No Escalada.

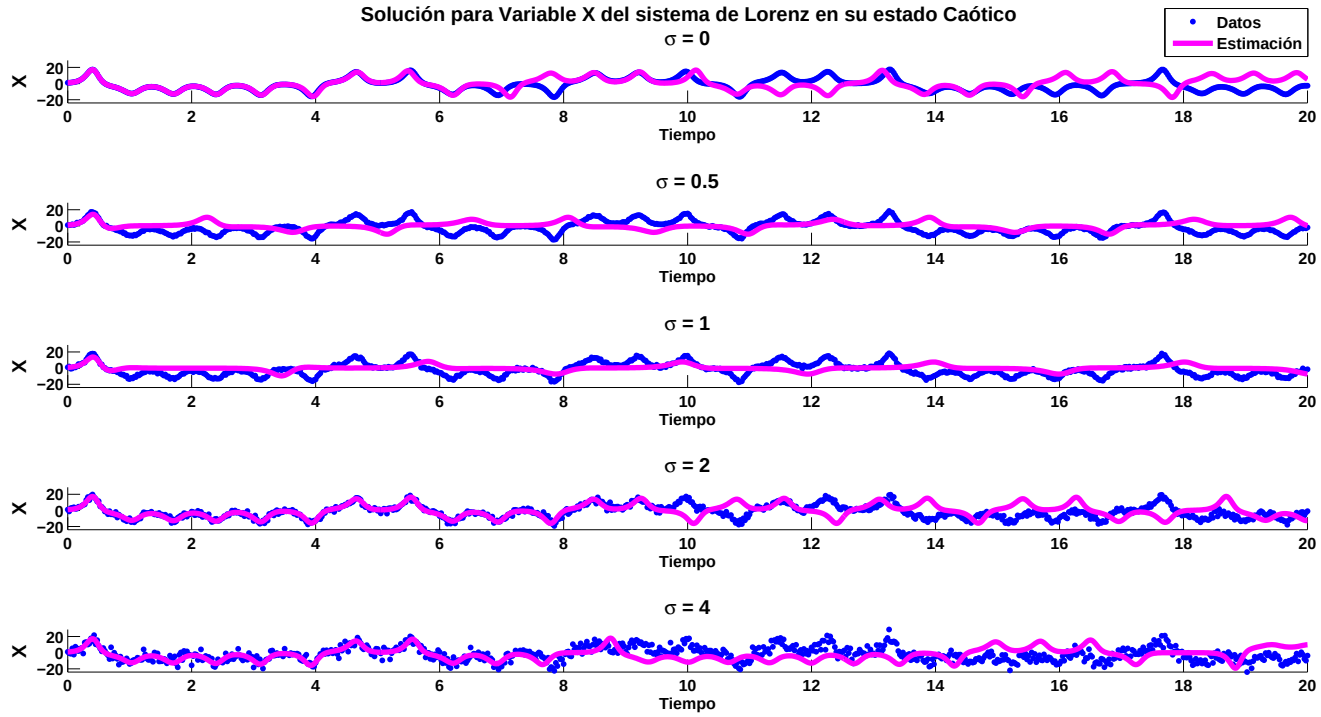

Figura 4.16: Por filas, la solución estimada (línea sólida magenta) y los datos con ruido (puntos azules) del sistema de Lorenz en su estado caótico a distintos niveles de ruido (σ =0, 0.5, 1, 2 y 4) para la variable X.

4.3.2. Soluciones

Las Figuras 4.16, 4.17 y 4.18, corresponden a las soluciones estimadas de las variables X, Y y Z, respectivamente, del sistema de Lorenz en su estado caótico a partir de los parámetros estimados por la moda de las distribuciones posteriores (Cuadro 4.2), del sistema de Lorenz en su estado caótico. Para las variables X, Y y Z para datos sin ruido ($\sigma = 0$) y para datos con ruido muestral de desviación estándar σ de 2 y 4, (filas 1, 4 y 5 de las Figuras 4.16, 4.17 y 4.18) se observa un desfase a mediano plazo de la solución estimada (línea sólida magenta) respecto a los datos (puntos azules), además de una pérdida de la trayectoria de los mismos. Sin embargo, al observar los planos fase (Figura 4.19) los gráficos correspondientes a estos niveles de ruido poseen una configuración muy similar, situación que no ocurre con los planos fase respectivos de las soluciones estimadas para los niveles de ruido σ de 0.5 y 1.

Para estas soluciones, se observa (Figuras 4.16, 4.17 y 4.18, filas 2 y 3) que hay un completo desfase entre los datos (puntos azules) y la solución estimada (línea sólida punteada). Se tiene que ésta última posee oscilaciones con una mayor longitud, lo anterior se refleja en el plano fase (Figura 4.19, paneles segunda fila) donde se observan trayectorias que oscilan en menor número alrededor de los puntos de equilibrio. Cabe señalar que en estos niveles de ruido se estimó el valor del parámetro **b** lejos de su valor real (Figuras 4.12 y 4.13).

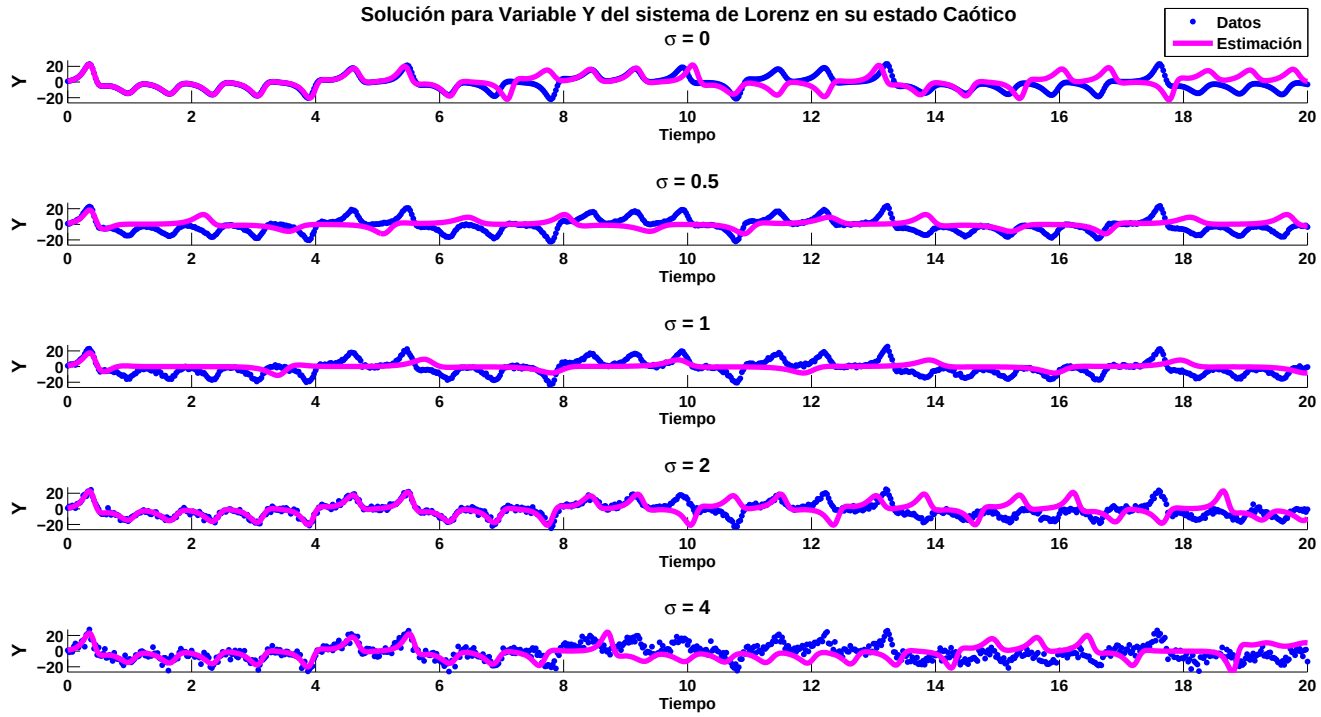

Figura 4.17: Por filas, la solución estimada (línea sólida magenta) y los datos con ruido (puntos azules) del sistema de Lorenz en su estado caótico a distintos niveles de ruido (σ =0, 0.5, 1, 2 y 4) para la variable Y.

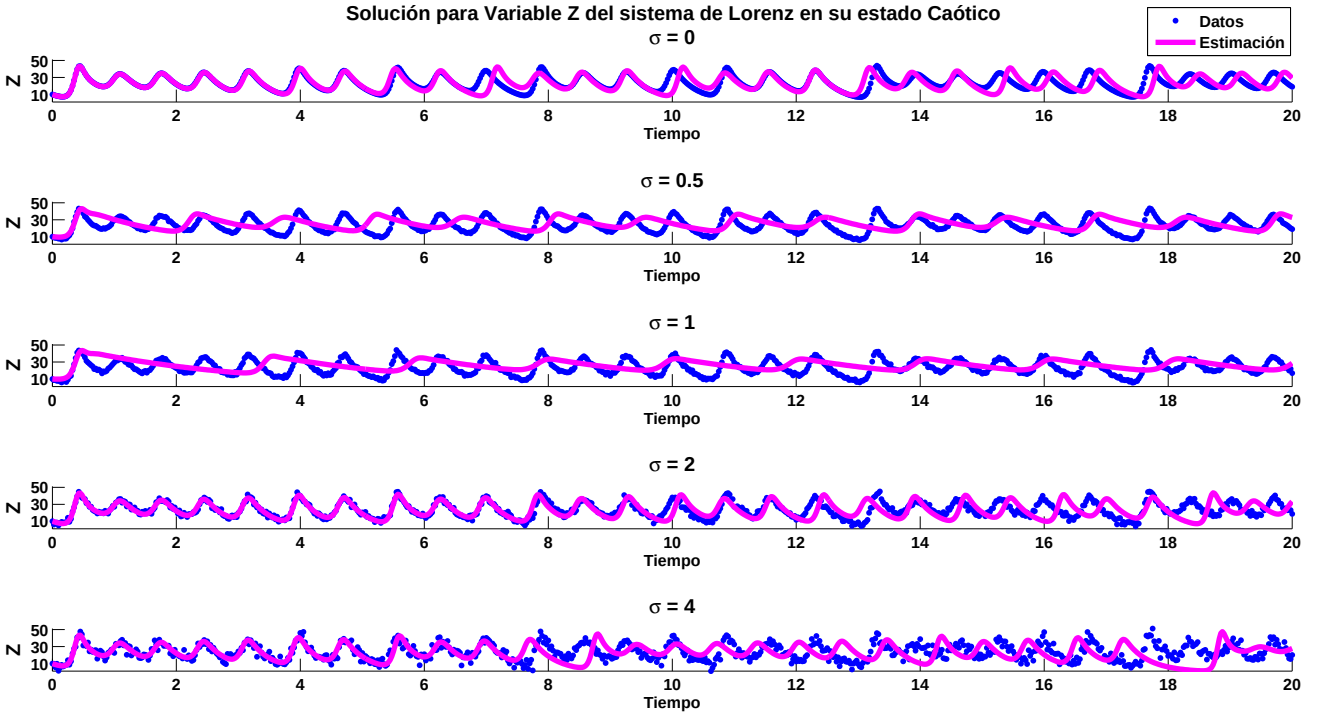

Figura 4.18: Por filas, la solución estimada (línea sólida magenta) y los datos con ruido (puntos azules) del sistema de Lorenz en su estado caótico a distintos niveles de ruido (σ =0, 0.5, 1, 2 y 4) para la variable Z.

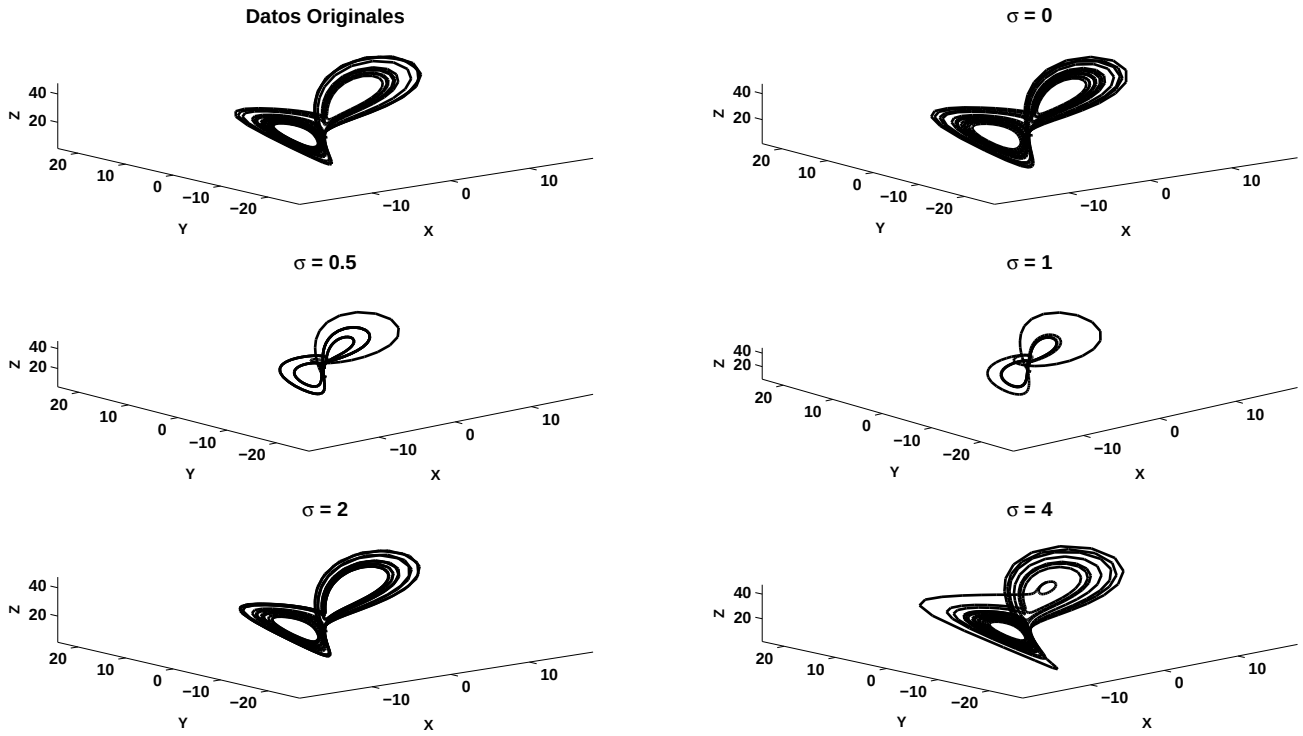

Figura 4.19: Planos fase de los datos originales (primer Figura esquina superior izquierda) y de la solución estimada (figuras restantes) del sistema de Lorenz (estado caótico) a distintos niveles de ruido.

Cuadro 4.3: Estimadores del valor del parámetro **b** a partir de la modas y moda secundaria, respectivamente, presentes en las distribuciones posteriores del parámetro para el sistema de Lorenz en su estado caótico para los niveles de ruido muestral de $\sigma = 0.5$ y 1, Figuras 4.12 y 4.13.

Parámetro **b**	Moda	Moda Secundaria
Ruido 0.5	0.6979239	2.7043751
Ruido 1	0.3659147	2.6760215

Como un anexo al análisis anterior se muestran en las Figuras 4.20, 4.21 y 4.22, las soluciones estimadas de las variables X, Y y Z, respectivamente, considerando ahora el valor del parámetro **b** dado por la moda secundaria de su distribución posterior, que resulta estar más cercano al valor real del parámetro (Figuras 4.12 y 4.13), Cuadro 4.3. Las soluciones estimadas presentan nuevamente un desfase a mediano plazo y una pérdida de la dinámica en el largo plazo, sin embargo, los planos fase muestran un configuración similar respecto a los datos originales y las demás soluciones a los distintos niveles de ruido, Figura 4.23, mostrando una recuperación de la tendencia del atractor.

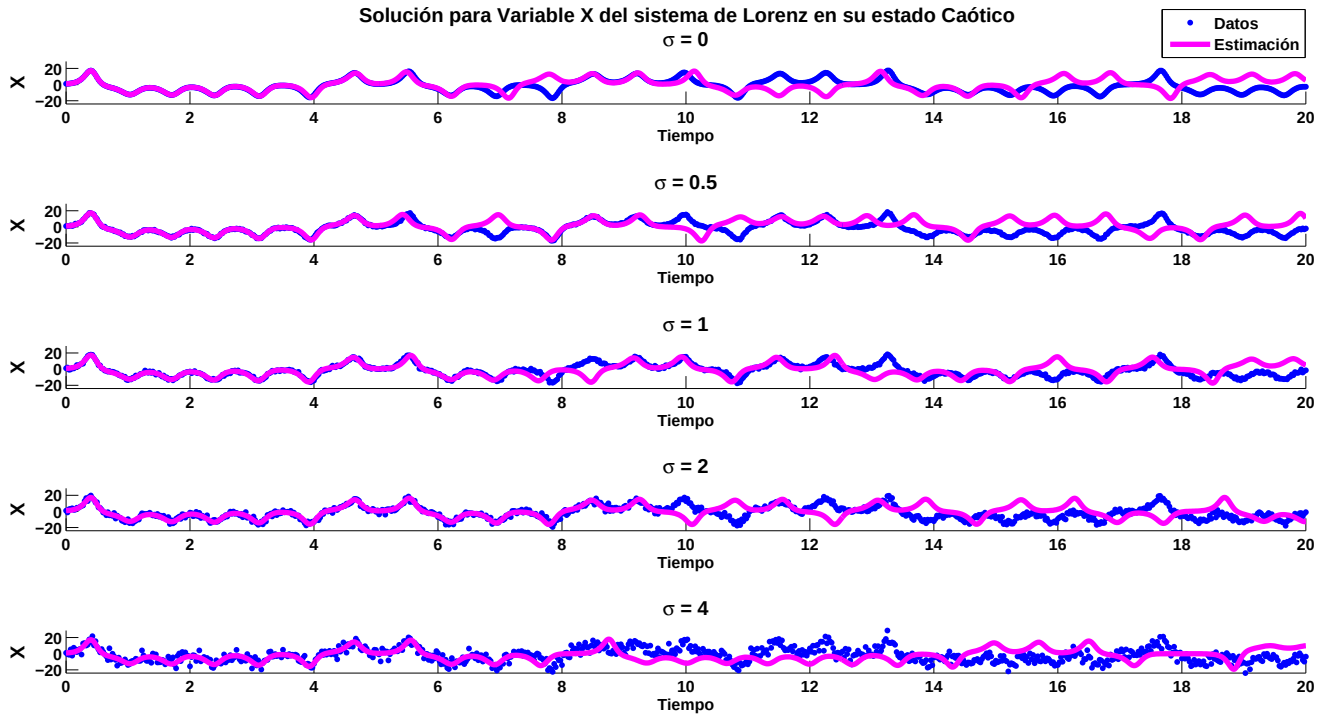

Figura 4.20: Por filas, la solución estimada (línea negra) y los datos con ruido (línea roja) del sistema de Lorenz (estado caótico) a distintos niveles de ruido para la variable X. Moda Secundaria.

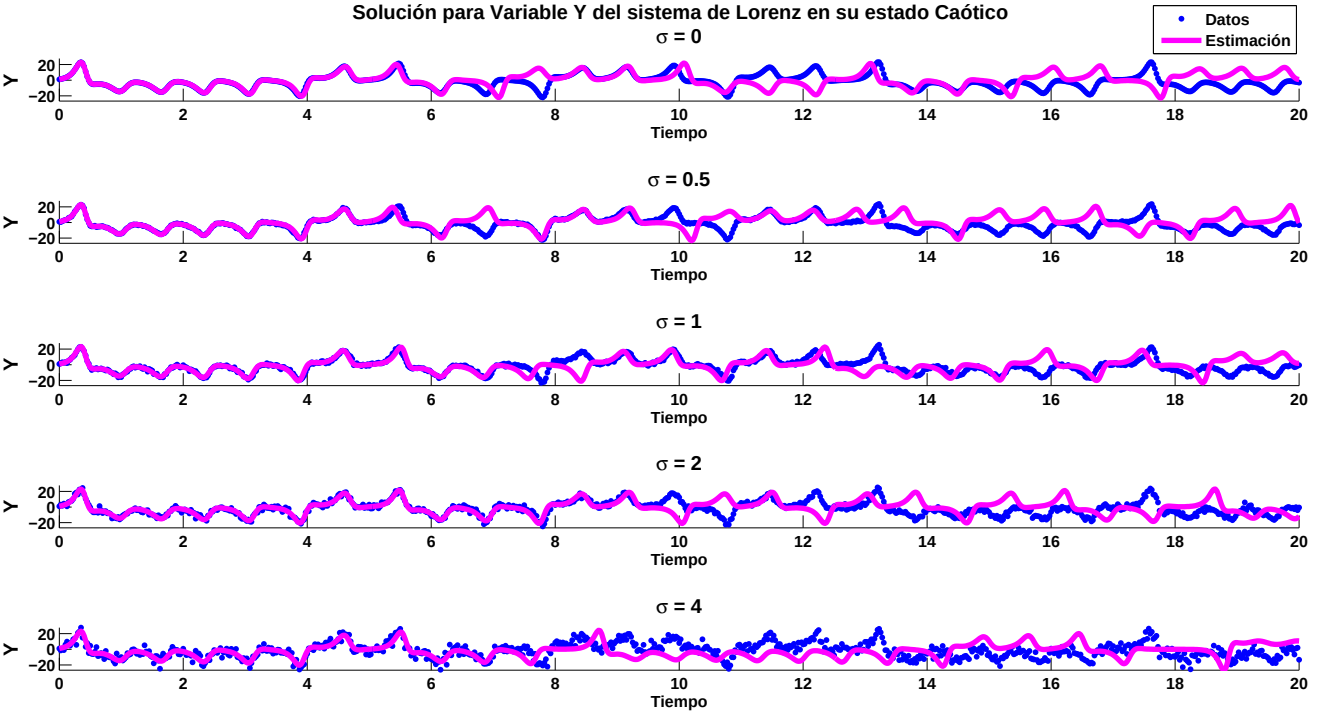

Figura 4.21: Por filas, la solución estimada (línea negra) y los datos con ruido (línea roja) del sistema de Lorenz (estado caótico) a distintos niveles de ruido para la variable Y. Moda Secundaria.

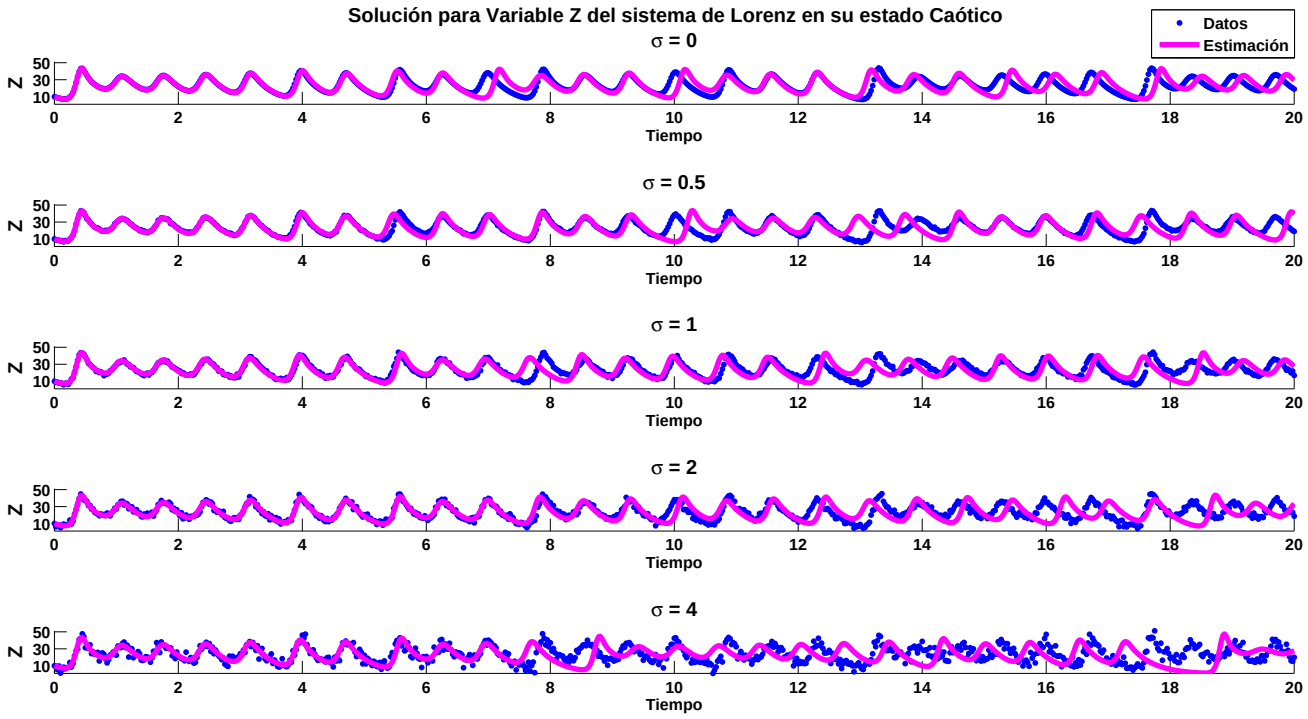

Figura 4.22: Por filas, la solución estimada (línea negra) y los datos con ruido (línea roja) del sistema de Lorenz (estado caótico) a distintos niveles de ruido para la variable Z. Moda Secundaria.

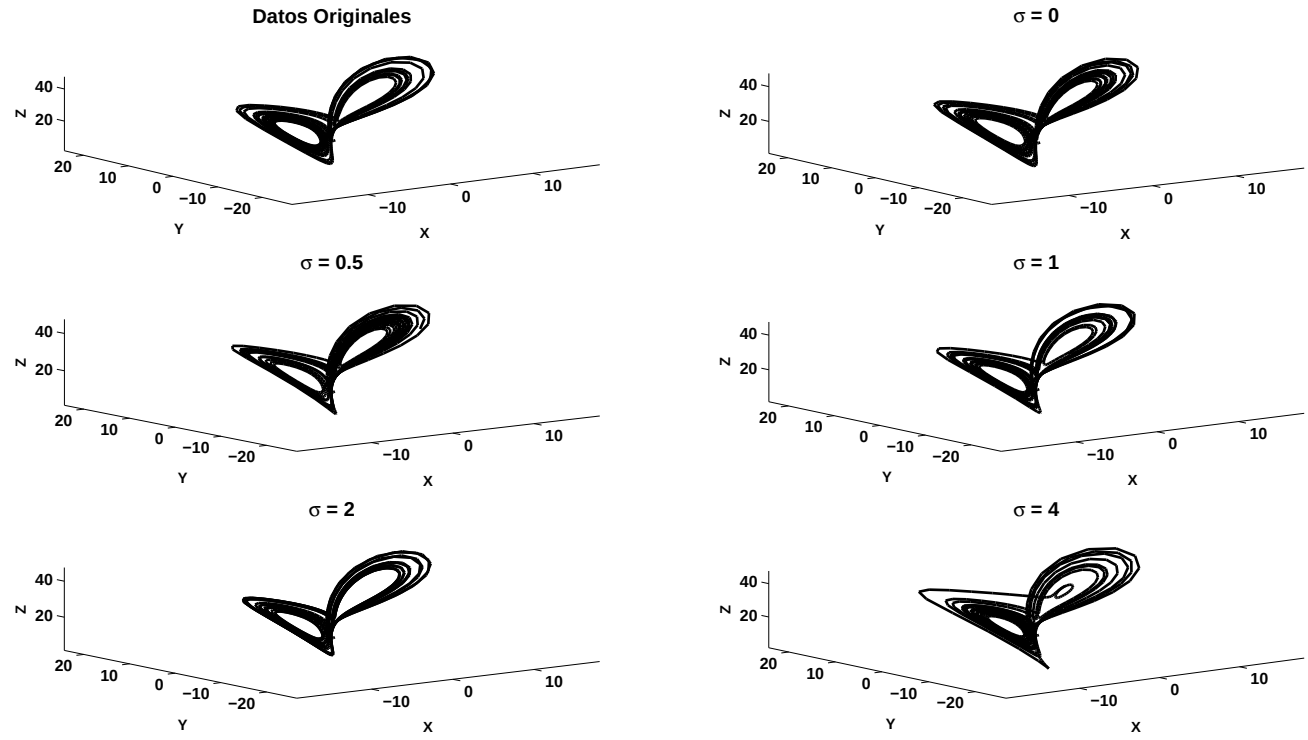

Figura 4.23: Planos fase de los datos originales (primer figura esquina superior izquierda) y de la solución estimada (figuras restantes) del sistema de Lorenz (estado caótico) a distintos niveles de ruido. Moda Secundaria.

4.4. Resultados y conclusiones

El sistema de Lorenz en lo referente a la estimación de sus parámetros y obtención de soluciones, presenta diversos problemas al enfrentarse al ruido muestral. En primer lugar, ante un estado no caótico, las distribuciones posteriores de los parámetros presentan regiones de alta densidad en sus extremos, particularmente en niveles de ruido medio ($\sigma = 0.5$ y 1), lo que se traduce en intervalos de probabilidad muy extensos, y dada la sensibilidad de las soluciones a los parámetros, los cataloga como intervalos poco informativos. En general se obtuvieron distribuciones unimodales lo que se traduce en un estimador más fiable y de hecho los estimadores se encontraron cerca del valor real del parámetro. Se destaca que la distribución correspondiente a nivel de ruido cero ($\sigma = 0$), si bien es unimodal, es la más dispersa de todas para cada uno de los parámetros.

En el caso del sistema en su estado caótico, las distribuciones posteriores presentaron una multimodalidad muy marcada, independientemente del ruido, sin embargo, cabe recalcar que a pesar de tal multimodalidad, las distribuciones fueron muy concentradas para niveles de ruido σ igual a 0, 2 y 4. Lo anterior contrasta con los intervalos de probabilidad extensos presentados en el caso del parámetro **b** (Figuras 4.12 y 4.13), donde se presentó una clara bimodalidad, lo que generó estimadores muy alejados del valor real. En lo referente a la dispersión de las distribuciones, es fácil observar que para el nivel de ruido más alto ($\sigma = 4$), se obtuvieron las distribuciones menos dispersas para todos los parámetros.

Para las soluciones en el estado no caótico, se destaca que para niveles de ruido $\sigma = 0$ y $\sigma = 4$, se presentaron desfases tempranos en la estimación respecto a los datos, además de una reducción en la amplitud de las oscilaciones, situación que no ocurre o se hace evidente, en niveles de ruido intermedio ($\sigma = 0.5$, 1 y 2), donde las soluciones estimadas resultan estar muy cercanas a los datos. En general, la tendencia de las soluciones (el atractor del sistema) se recupera sin mayor problema independientemente del ruido muestral, como lo muestra su plano fase (Figura 4.9).

Finalmente, para las soluciones correspondientes al estado caótico, se presenta un desfase a mediano plazo para las soluciones correspondientes a los niveles de ruido $\sigma = 0$, 2 y 4. Nuevamente, tal desfase es seguido de una pérdida completa de las soluciones respecto a los datos y no se presenta una reducción evidente en la amplitud de las soluciones como lo ocurrido en el estado no caótico. Para las soluciones correspondientes para datos con ruido de desviación estándar $\sigma = 0.5$ y 1, ocurrió una pérdida completa de la solución respecto a los datos. La amplitud de las oscilaciones de redujo y ocurrió un aumento en la longitud de las mismas, y es evidente la diferencia de estas soluciones en sus planos fase respectivos (Figura 4.19, segunda fila).

Una estimación a largo plazo, independientemente del estado del sistema, caótico o no caótico, resulta ser difícil o imposible, como lo mostró la estimación para el sistema en su estado caótico, no así la estimación del atractor del sistema, que se recupera de manera razonable, como se observa en los planos fases respectivos para cada estado (Figuras 4.9 y 4.19; no caótico y caótico, respectivamente) donde los mismos guardan cierta similitud respecto al plano fase de los datos sin ruido. Lo anterior implica que la tendencia del atractor puede estimarse sin mayor problema, salvo las dificultades ya mencionadas del parámetro **b** en niveles de ruido con desviación estándar $\sigma = 0.5$ y 1.

En el artículo, “Effect of sample noise on the parameter estimation of complex dynamic systems”, de Chávez y Castaño (2014), se realizó la estimación de parámetros del sistema de Lorenz en su estado no caótico, sin embargo se añadió ruido a las condiciones iniciales de las variables, sabiendo que el sistema es sensible al cambio en las mismas. Se observaron resultados similares al análisis ya expuesto. Por un lado, la fiabilidad de la estimación dependerá del horizonte de estimación deseado. En el corto plazo, la estimación resulta ser más precisa (se presentan desfases a partir del mediano plazo de las soluciones estimadas), mientras que al largo plazo se presentan problemas de recuperación de los datos. Finalmente, en el mismo artículo, se redujo la complejidad de la dinámica al considerar una subregión de la misma (subconjunto de datos de la dinámica), resultando en una mejor estimación, lo que se traduce en una mejor estimación por intervalos que en una dinámica completa. Nuevamente, la interpretación de la tendencia a largo plazo no se vé afectada pues los efectos del ruido sobre el atractor del sistema no son considerables.

Al contrastar ambos estados del sistema de Lorenz, puede concluirse que el factor determinante de la estimación no es el ruido muestral y mucho menos algún problema de especificación, más bien la complejidad del fenómeno, pues como se ha observado, es la complejidad la que se ve reflejada en las distribuciones posteriores y las soluciones estimadas, aún de manera más marcada que el ruido mismo.

5. Aplicación: Biorreactor

Introducción

El proceso de modelado, como se ha visto, se enfrenta a diversos problemas y en su mayoría son debidos a la falta de información disponible del fenómeno en estudio. Los problemas de especificación del modelo pueden deberse a la omisión o inclusión de variables importantes o innecesarias, respectivamente. Aún si se conocieran todas las variables inmiscuidas, queda el problema de saber cómo utilizarlas en una correcta forma funcional; para así recuperar el fenómeno en cuestión. El caso de estudio que se presenta en seguida muestra un proceso de modelación de un fenómeno químico, el cual se enfrenta a problemas de especificación, resultando en una serie de modelos y modificaciones que intentan recuperar los datos experimentales. Tales problemas de especificación se ven reflejados en las distribuciones posteriores de los parámetros estimados a través de métodos Bayesiano.

Con apoyo de la Facultad de Química se obtuvo acceso a un conjunto de datos experimentales correspondiente a un proceso químico llevado a cabo dentro de un biorreactor. El modelo que intenta explicar las reacciones dentro del biorreactor consta de un sistema de tres ecuaciones ordinarias de primer orden con tres variables, $X_1 =$ Biomasa, $X_2 =$ Ácido Láctico y $X_3 =$ Lactosa; además de un conjunto adicional de parámetros $\boldsymbol{\theta}$, cuya dimensión dependerá del modelo propuesto que se esté utilizando. De manera general, el sistema se representa por la expresión siguiente:

$$\begin{aligned} \frac{dX_1}{dt} &= p\left(X_1, X_2, X_3; \boldsymbol{\theta}\right) \\ \frac{dX_2}{dt} &= q\left(X_1, X_2, X_3; \boldsymbol{\theta}\right) \\ \frac{dX_3}{dt} &= r\left(X_1, X_2, X_3; \boldsymbol{\theta}\right) \end{aligned}$$

Para esta aplicación se poseen dos conjunto de datos, cada uno con tres réplicas experimentales. Ambos conjuntos de datos se presentan en las Figuras 5.1 y 5.2. Para el proceso de modelación-estimación se utilizó el algoritmo population-based MCMC. Se propusieron varios modelos y modificaciones de los mismos, sin embargo aquí sólo se muestran aquellos modelos representativos del proceso de estimación.

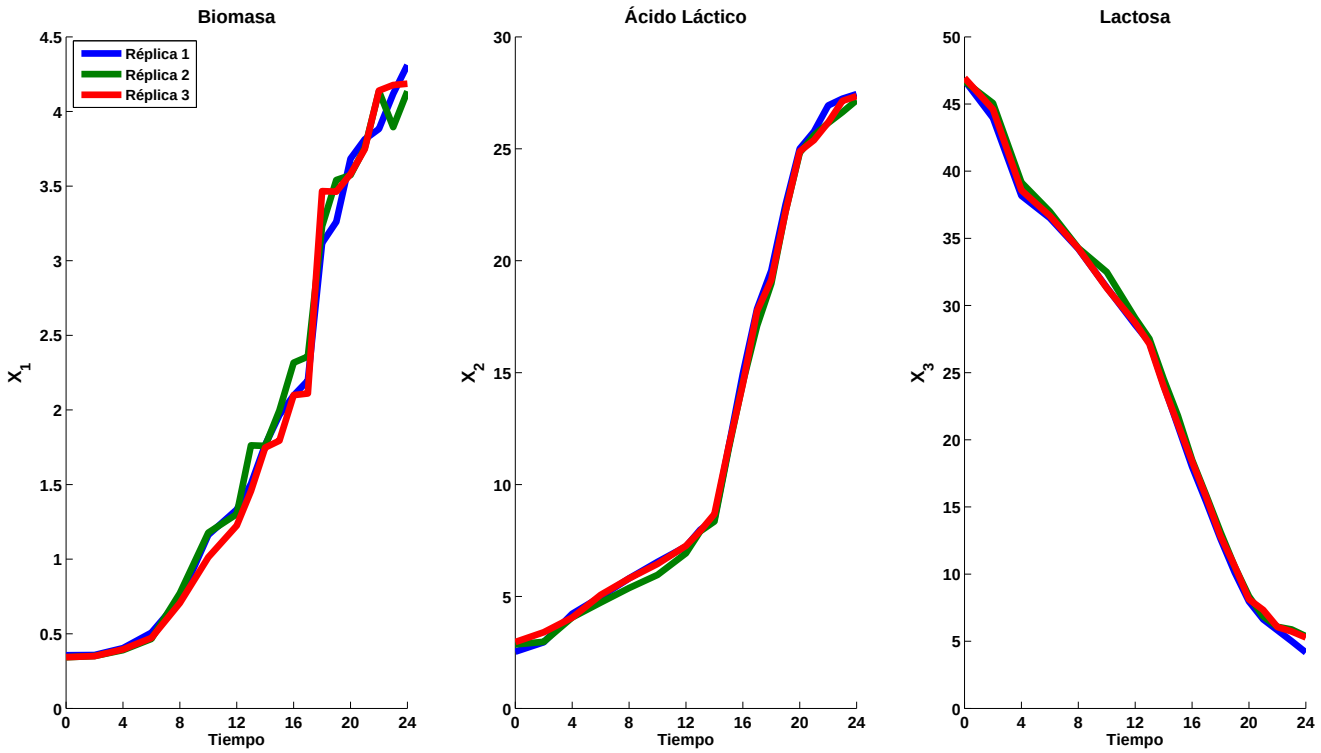

Figura 5.1: Dinámicas observadas junto con las réplicas para las variables Biomasa, Ácido Láctico y Lactosa, respectivamente, para el primer conjunto de datos.

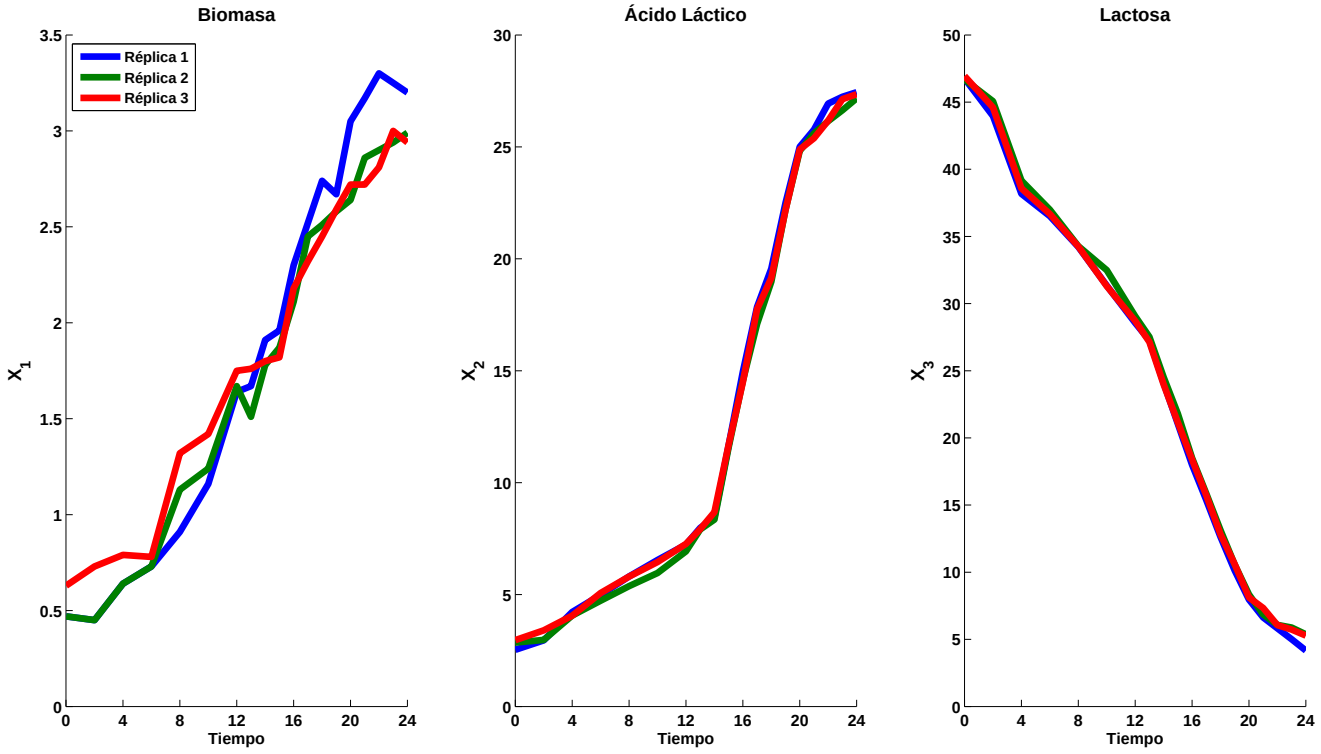

Figura 5.2: Dinámicas observadas junto con las réplicas para las variables Biomasa, Ácido Láctico y Lactosa, respectivamente, para el segundo conjunto de datos.

5.1. Primer modelo propuesto

El primer modelo propuesto utilizado para explicar los datos experimentales consta de siete parámetros y el sistema está dado por la expresión (5.1). Los parámetros iniciales corresponden a la expresión (5.2); es menester señalar que se estimó el parámetro $Xmax$, sin embargo, puesto que representa un valor máximo obtenido experimentalmente, debiera considerarse como un parámetro fijo.

$$\begin{aligned} \frac{dX_1}{dt} &= \frac{umax \cdot X_3}{ks + X_3} \cdot \left[1 - \frac{X_1}{Xmax}\right] \\ \frac{dX_2}{dt} &= a \cdot \frac{dX_1}{dt} + b \cdot X_1 \\ \frac{dX_3}{dt} &= -\frac{1}{Yps} \cdot \frac{dX_2}{dt} - ms \cdot X_1 \end{aligned} \tag{5.1}$$

$$\boldsymbol{\theta}_0 = \begin{bmatrix} umax \\ ks \\ Xmax \\ a \\ b \\ Yps \\ ms \end{bmatrix} = \begin{bmatrix} 0.154 \\ 7.22 \\ 2.842 \\ 7.31 \\ 0.659 \\ 0.57 \\ 0.865 \end{bmatrix} \tag{5.2}$$

Estimación y Solución

Los parámetros estimados por la moda de la distribución posterior (Figura 5.4) se muestran en el Cuadro 5.1. La Figura 5.3 muestra las soluciones estimadas para las tres variables. En primer lugar, para la variable biomasa (X_1, primer panel), se observa que la solución estimada comienza con una buena recuperación de los datos en una fase temprana para posteriormente comenzar a subestimar la dinámica de los datos. En el caso de la variable ácido láctico (X_2, segundo panel), se observa que la solución estimada recupera la tendencia creciente de los datos hasta el término de la dinámica, sin embargo, en el mediano plazo no se recuperan los datos pues los subestima y sobreestima después del inicio y antes del término de la dinámica, respectivamente. Finalmente, para la tercera variable, lactosa (X_3, tercer panel), los datos se recuperan de manera más satisfactoria que en las otras dos variables a lo largo de todo el periodo de tiempo.

La Figura 5.4 muestra las distribuciones posteriores de los parámetros. Se observan

distribuciones claramente sesgadas y en algunos casos una aparente multimodalidad (parámetros $Xmax$, a y ms, filas 3, 4 y 7, respectivamente). El sesgo en las distribuciones se debe a la presencia de regiones de alta probabilidad en las colas de las distribuciones lo que resulta en intervalos de probabilidad muy amplios.

Cuadro 5.1: Parámetros estimados por la moda de la distribución posterior (Figura 5.4) para el modelo (5.1).

Parámetro	$Umax$	ks	$Xmax$	a	b	Yps	ms
Estimación	0.0821	7.126	4.7311	4.0446	0.692	0.5419	0.1139

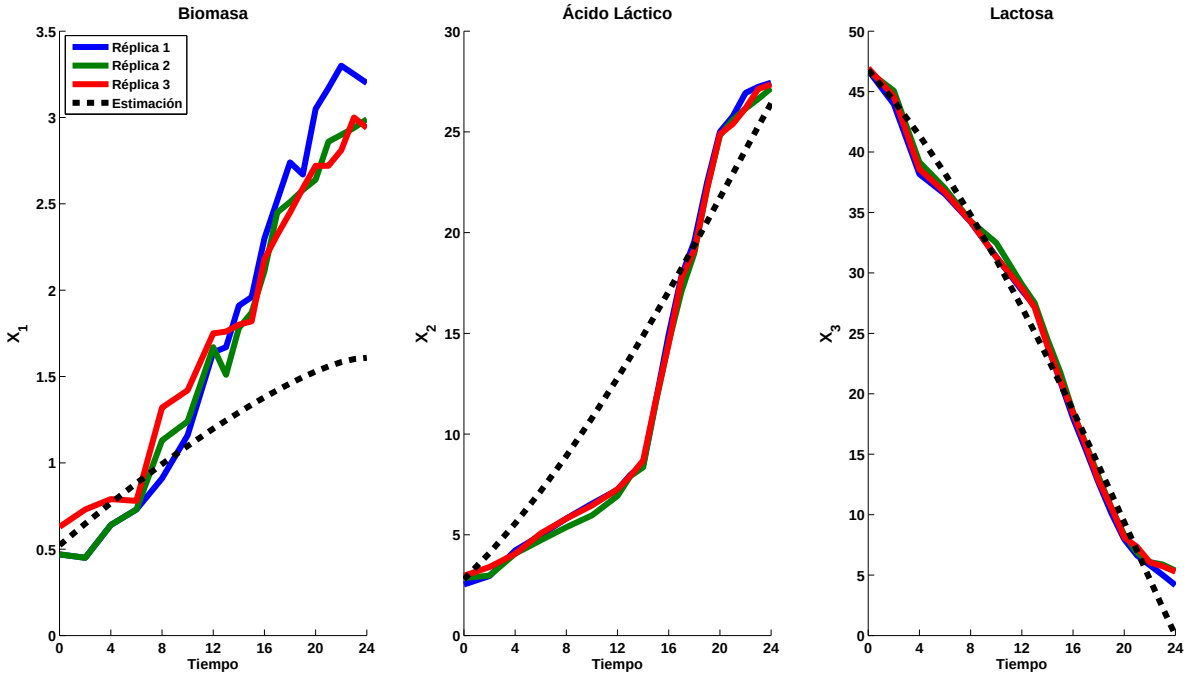

Figura 5.3: Comparativa entre los datos experimentales (líneas de color) y las soluciones estimadas (líneas negras punteadas) para el modelo (5.1).

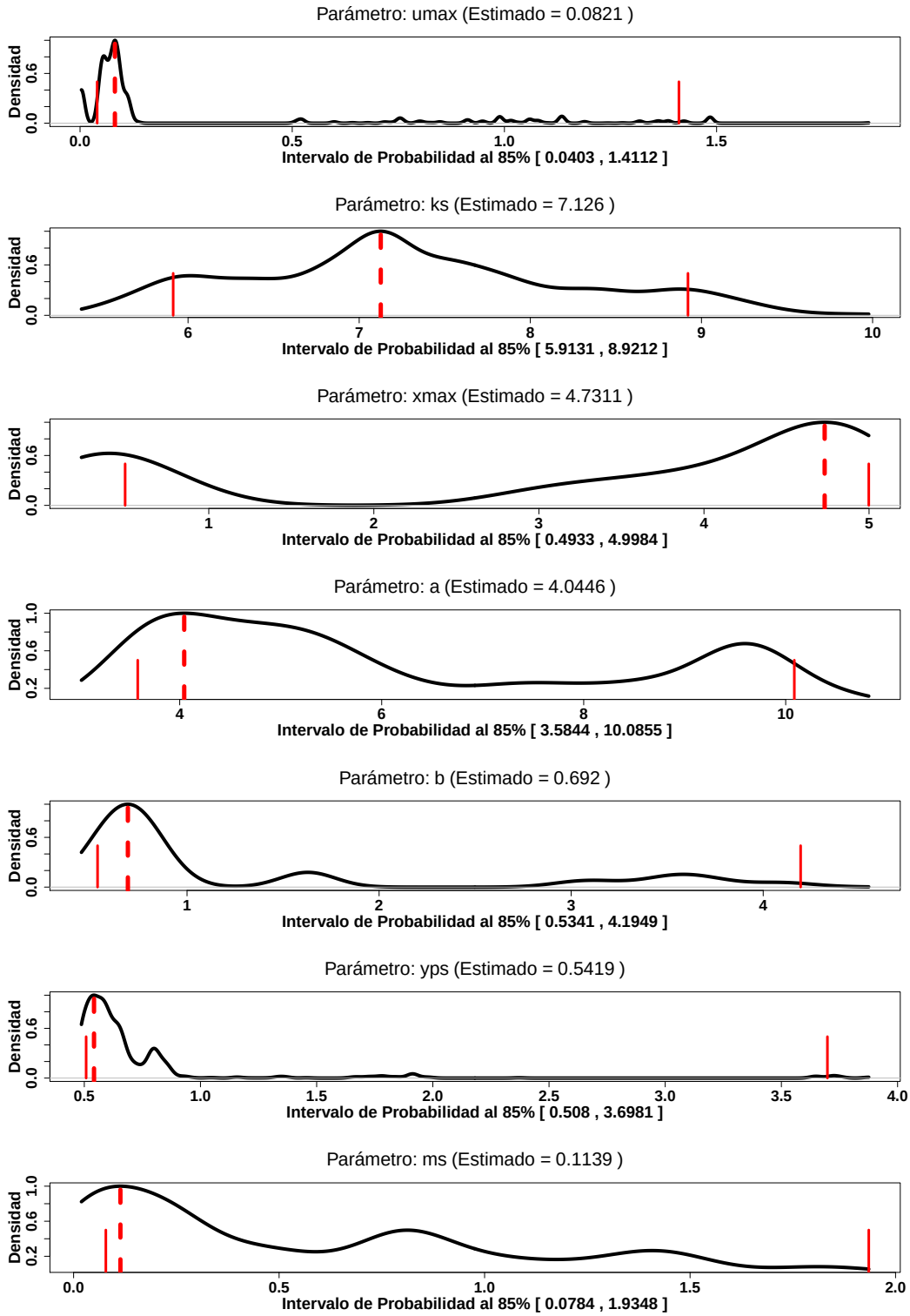

Figura 5.4: Por filas, las distribuciones posteriores de los parámetros del modelo (5.1). La línea roja cortada corresponde al valor estimado por la moda, las líneas rojas sólidas corresponden a los extremos del intervalo de probabilidad al 85 %.

5.2. Segunda propuesta

Dados los problemas surgidos en el primer modelo respecto a la estimación de las variables biomasa y ácido láctico, se procedió a un segundo enfoque correspondiente al sistema de ecuaciones (5.3), el cual consistente ahora de 5 parámetros. Este modelo considera a la razón de cambio de la biomasa ($\frac{dX_1}{dt}$) como constante, dado el aparente comportamiento lineal de los datos, además de que la razón de cambio del ácido láctico ($\frac{dX_2}{dt}$) pasa a ser una función lineal de la biomasa (X_1).

$$\begin{aligned} \frac{dX_1}{dt} &= v \\ \frac{dX_2}{dt} &= a \cdot v + b \cdot X_1 \\ \frac{dX_3}{dt} &= -\frac{1}{Yps} \cdot \frac{dX_2}{dt} - ms \cdot X_1 \end{aligned} \tag{5.3}$$

Los parámetros iniciales se muestran en la expresión (5.4).

$$\boldsymbol{\theta}_0 = \begin{bmatrix} v \\ a \\ b \\ Yps \\ ms \end{bmatrix} = \begin{bmatrix} 0.018 \\ 7.31 \\ 0.659 \\ 0.57 \\ 0.865 \end{bmatrix} \tag{5.4}$$

Estimación y Solución

La Figura 5.5 muestran las soluciones estimadas a partir de los parámetros del Cuadro 5.2, obtenidos a partir de la moda de las distribuciones posteriores mostradas en la Figura 5.6. En esta segunda estimación sigue habiendo problemas en la recuperación de los datos experimentales. Al igual que con el primer enfoque, se recupera la tendencia de la variable ácido láctico, así como se recupera razonablemente la dinámica de la lactosa, sin embargo el supuesto del comportamiento lineal en el caso de biomasa no recuperan los datos experimentales.

Por el lado de las distribuciones posteriores (Figura 5.6), se obtuvieron distribuciones más concentradas alrededor de la moda. Existe cierta mejoría respecto a la estimación anterior, sin embargo siguen habiendo concentraciones en los extremos y a lo largo de las mismas que arrojan intervalos de confianza muy extensos y por tanto poco informativos.

Cuadro 5.2: Parámetros estimados por la moda de la distribución posterior (Figura 5.6) para el modelo (5.3).

Parámetro	v	a	b	Yps	ms
Estimación	0.098	7.0962	0.1914	0.4607	0.0541

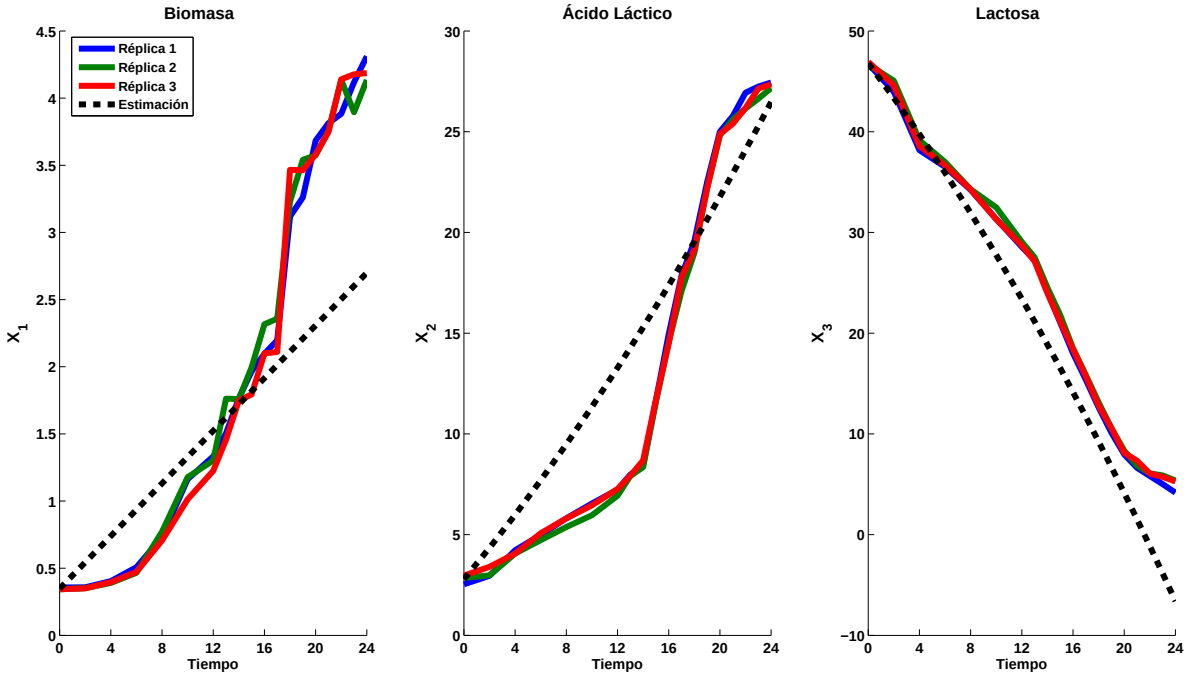

Figura 5.5: Comparativa entre los datos experimentales (líneas de color) y las soluciones estimadas (líneas negras punteadas) para el modelo (5.3).

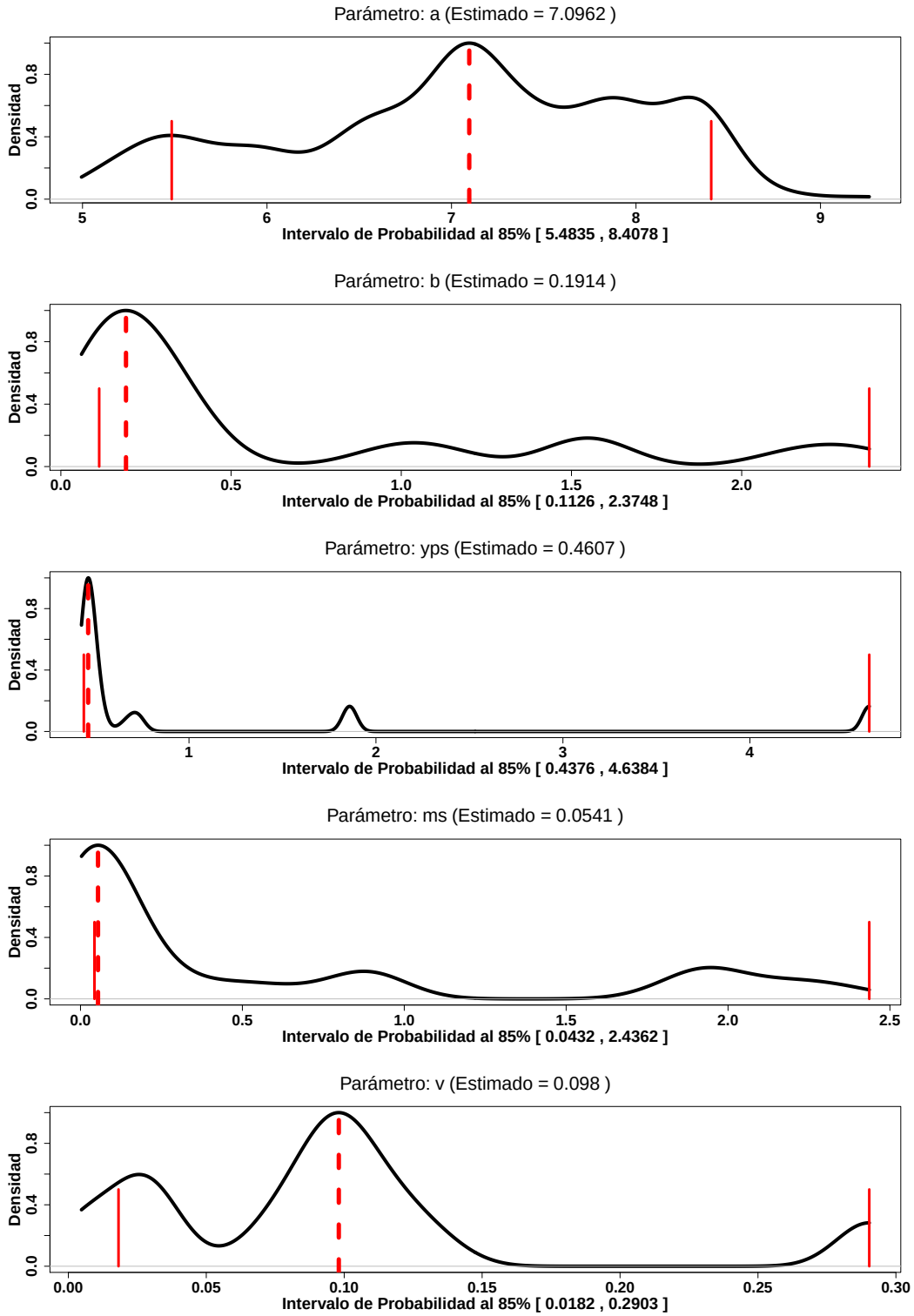

Figura 5.6: Por filas, las distribuciones posteriores de los parámetros del modelo (5.3). La línea roja cortada corresponde al valor estimado por la moda, las líneas rojas sólidas corresponden a los extremos del intervalo de probabilidad al 85 %.

5.3. Modelo logístico

Siguiendo en la búsqueda de un modelo que recupere la dinámica experimental se propone un nuevo modelo que se denominará "logístico" con la configuración dada por la expresión (5.5).

$$\begin{aligned} \frac{dX_1}{dt} &= \frac{k \cdot Xmax \cdot \exp\left(k \cdot (tc - t)\right)}{\left[1 + \exp\left(k \cdot (tc - t)\right)\right]^2} \\ \frac{dX_2}{dt} &= a \cdot \frac{dX_1}{dt} + b \cdot X_1 \\ \frac{dX_3}{dt} &= -\frac{1}{Yps} \cdot \frac{dX_2}{dt} - ms \cdot X_1 \end{aligned} \tag{5.5}$$

Los parámetros iniciales se muestran en la expresión siguiente:

$$\boldsymbol{\theta}_0 = \begin{bmatrix} k \\ tc \\ Xmax \\ a \\ b \\ Yps \\ ms \end{bmatrix} = \begin{bmatrix} 0.1511 \\ 14.4934 \\ 3.9282 \\ 7.31 \\ 0.659 \\ 0.57 \\ 0.865 \end{bmatrix}$$

Estimación y Solución

La Figura 5.7 muestran las soluciones estimadas versus los datos experimentales a partir de los parámetros del Cuadro 5.3. Las soluciones estimadas para el modelo logístico (5.5), Figura 5.7, muestran una clara mejoría en la estimación respecto a modelos anteriores para el caso de los datos de biomasa y lactosa, sin embargo, siguen habiendo problemas en la solución estimada para el ácido láctico que sólo recupera la tendencia creciente de los datos, no así la dinámica en el mediano plazo.

Observando las distribuciones posteriores correspondientes se tiene presencia de distribuciones multimodalidades, altamente dispersas, y por tanto, con intervalos de probabilidad muy extensos y poco informativos.

Cuadro 5.3: Por filas los parámetros estimados por la moda de las distribuciones posteriores (Figuras 5.8) para los modelos (5.5).

Parámetro	k	tc	$Xmax$	a	b	Yps	ms
Estimación	0.8583	15.3519	1.4062	7.9881	0.6611	8.4188	2.8039

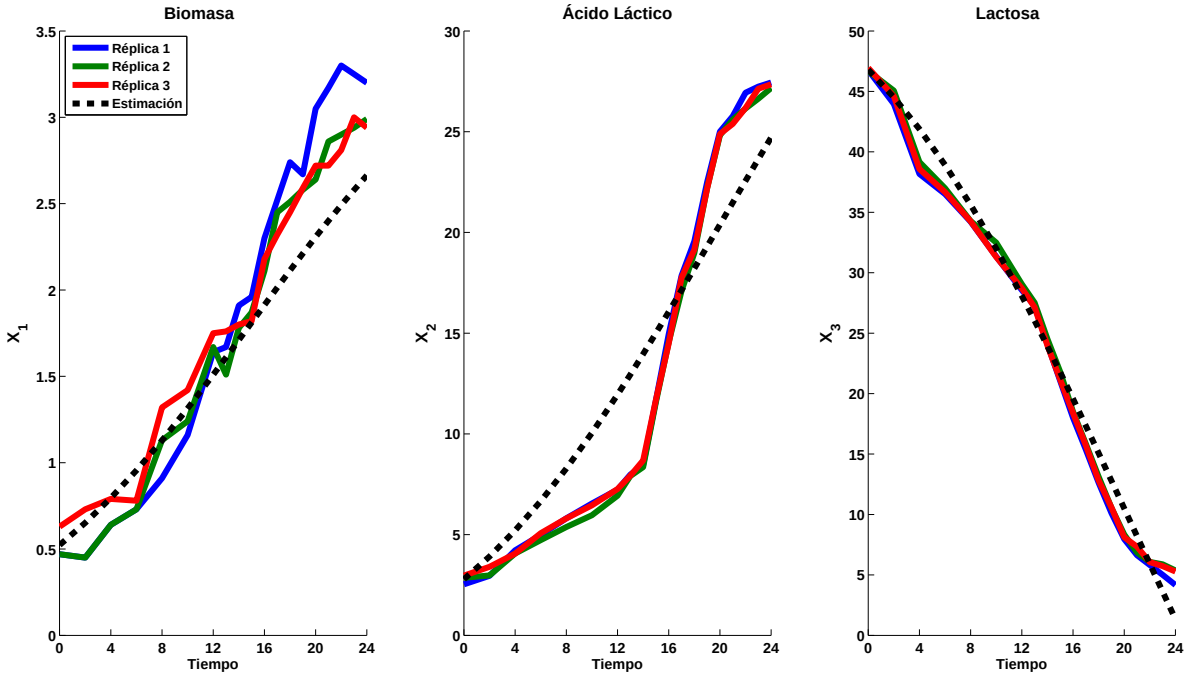

Figura 5.7: Comparativa entre los datos experimentales (líneas de color) y las soluciones estimadas (líneas negras punteadas) para el modelo (5.5).

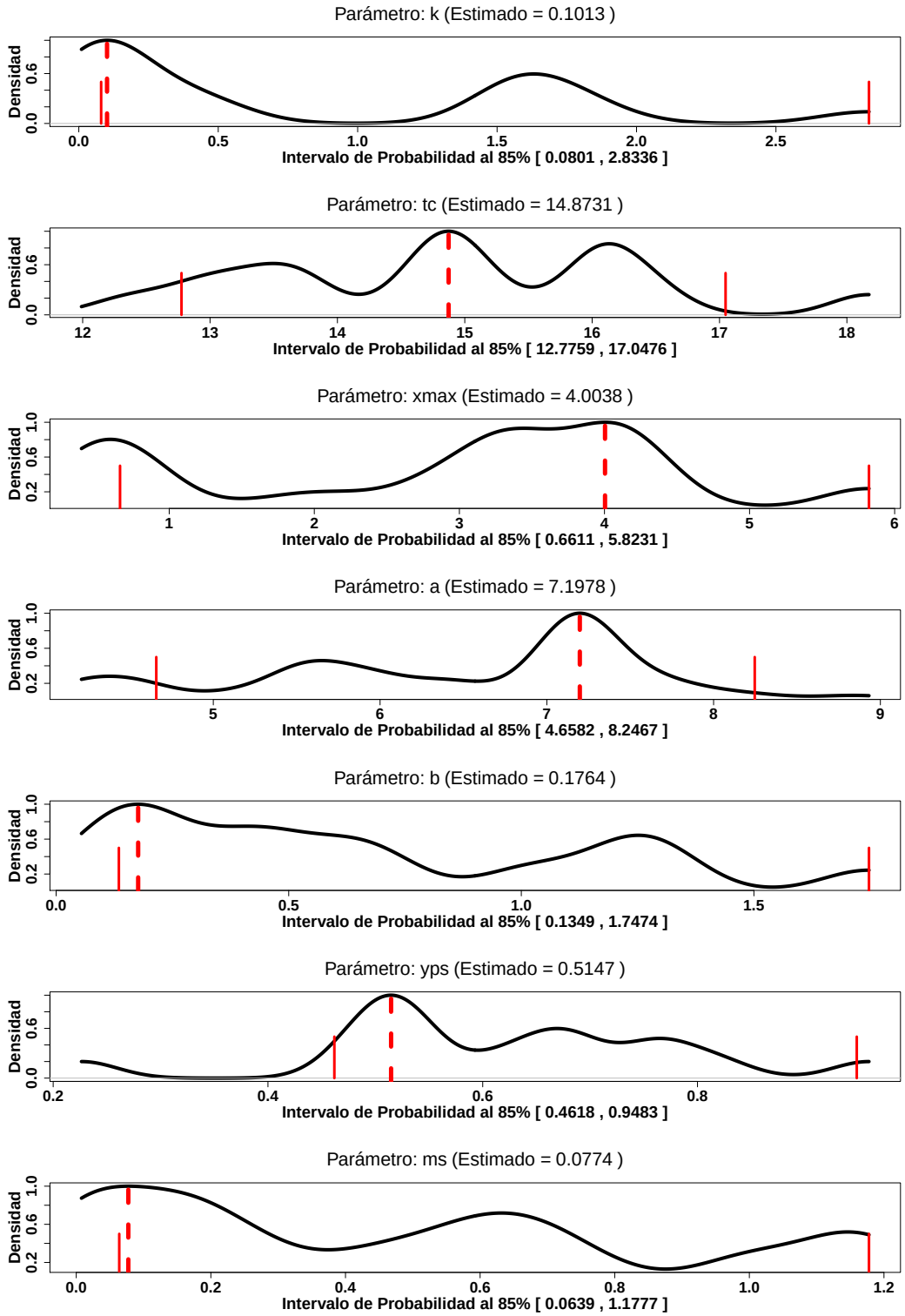

Figura 5.8: Por filas, las distribuciones posteriores de los parámetros del modelo (5.5). La línea roja cortada corresponde al valor estimado por la moda, las líneas rojas sólidas corresponden a los extremos del intervalo de probabilidad al 85 %.

5.4. Modelo logístico reducido

Siguiendo con el modelo logístico (5.5), fue modificado en su tercera ecuación eliminando el término –$ms \cdot X_1$, obteniendo el modelo en la expresión (5.6). Nuevamente el valor de $Xmax$ se consideró como un parámetro fijo.

$$\begin{aligned} \frac{dX_1}{dt} &= \frac{k \cdot Xmax \cdot \exp\left(k \cdot (tc - t)\right)}{\left[1 + \exp\left(k \cdot (tc - t)\right)\right]^2} \\ \frac{dX_2}{dt} &= a \cdot \frac{dX_1}{dt} + b \cdot X_1 \\ \frac{dX_3}{dt} &= -\frac{1}{Yps} \cdot \frac{dX_2}{dt} \end{aligned} \tag{5.6}$$

Estimación y Solución

El Cuadro 5.4 muestra los parámetros estimados del modelo (5.6) a partir de las distribuciones posteriores mostradas en la Figuras 5.10. Se observa un mejora en las soluciones estimadas respecto a la anterior estimación para la variable biomasa, Figuras 5.9. Sigue habiendo problemas en la estimación de la dinámica de la variable ácido láctico, X_2. Salvo las distribuciones de los parámetros tc y a; se tienen distribuciones unimodales muy concentradas alrededor de la moda de la distribución, sin embargo, existen regiones de densidad alejadas de la moda de las distribuciones lo que resulta en intervalos de probabilidad muy amplios.

Cuadro 5.4: Por filas los parámetros estimados por la moda de las distribuciones posteriores (Figuras 5.10 para los modelos (5.6).

Parámetro	k	tc	$Xmax$	a	b	Yps
Estimación	0.1529	14.367	3.9282	7.1198	0.0601	0.509

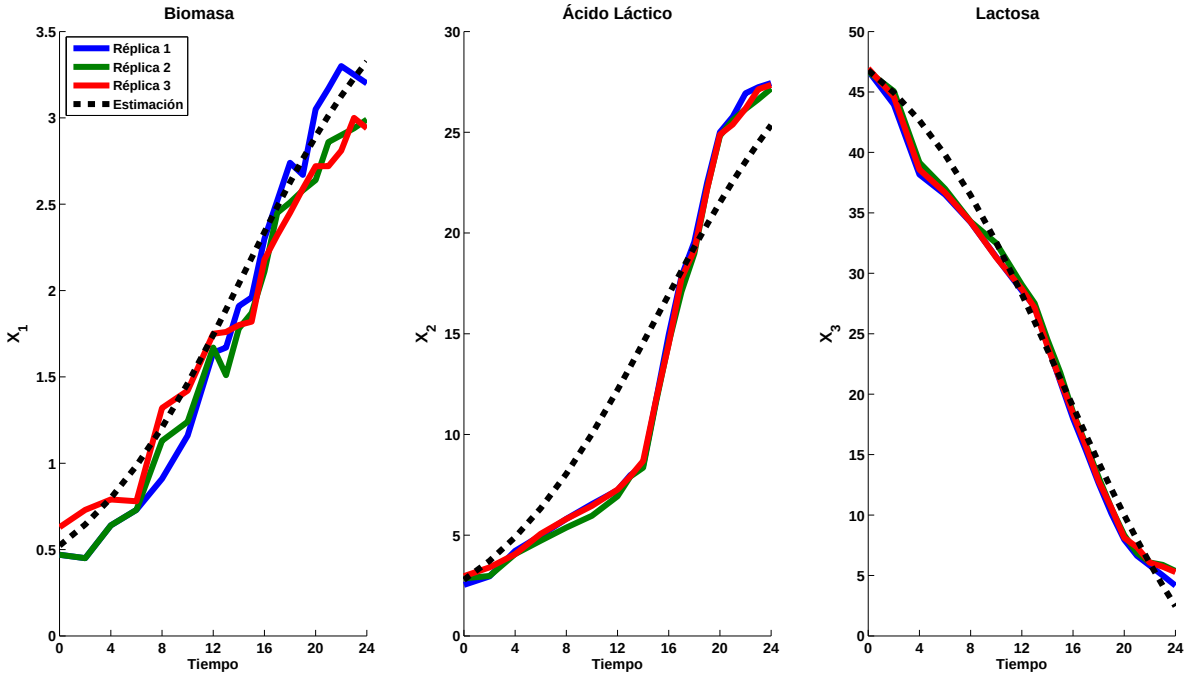

Figura 5.9: Comparativa entre los datos experimentales (líneas de color) y las soluciones estimadas (líneas negras punteadas) para el modelo (5.6).

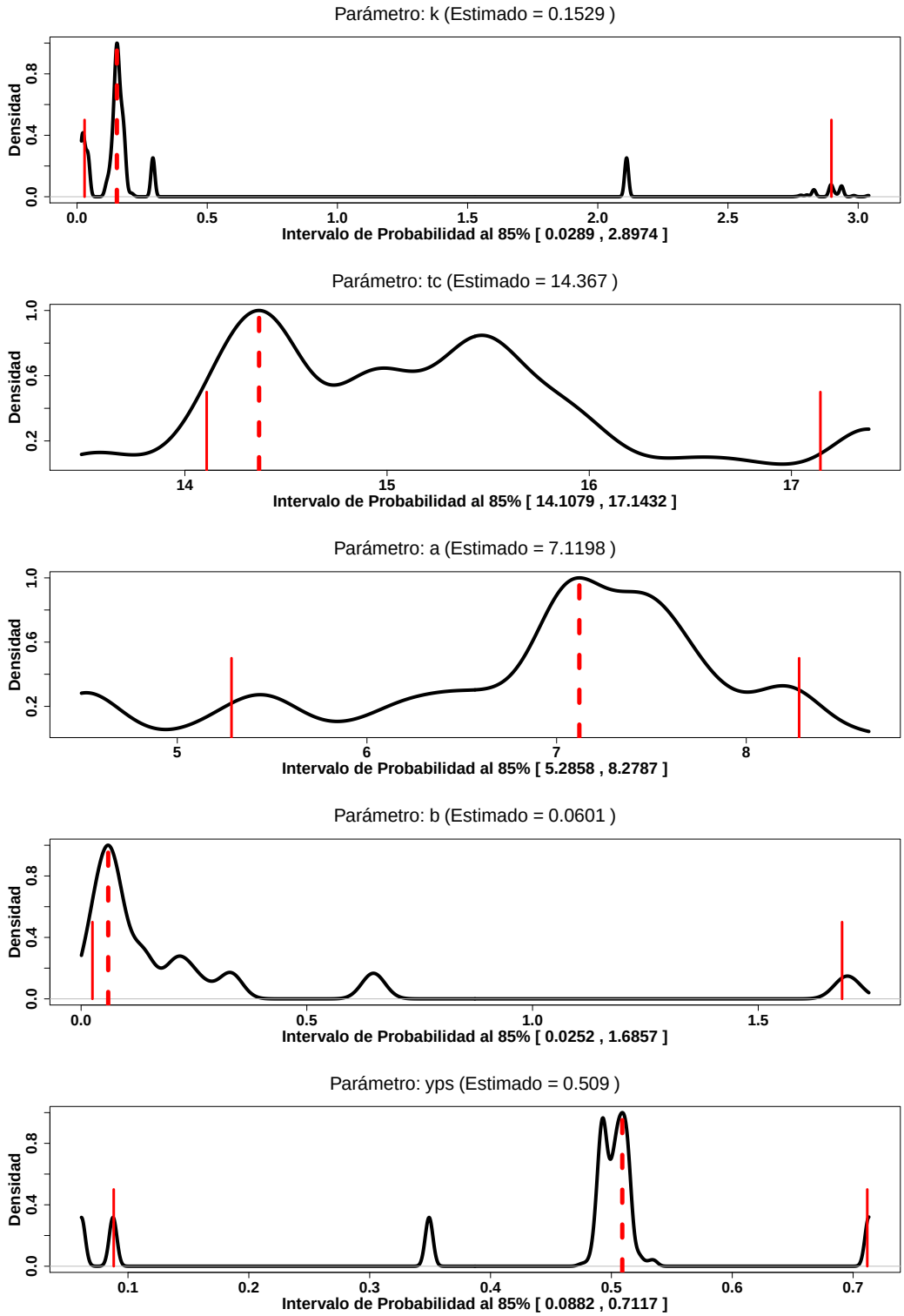

Figura 5.10: Por filas, las distribuciones posteriores de los parámetros del modelo (5.6). La línea roja cortada corresponde al valor estimado por la moda, las líneas rojas sólidas corresponden a los extremos del intervalo de probabilidad al 85 %.

Modelo Logístico en un Subconjunto de Tiempo

Se prosiguió con una nueva estimación considerando nuevamente el modelo logístico original, expresión (5.5), ahora sobre un subconjunto de datos, que corresponde al periodo de tiempo $[0, 17]$. Tal elección corresponde a el tiempo donde la variable ácido láctico comienza a cesar en su crecimiento.

Estimación y Solución

El Cuadro 5.5 muestra los parámetros estimados para esta nueva aproximación. Salvo la solución estimada (Figura 5.11) para la variable biomasa, que permanece muy alejada de los datos experimentales, las estimaciones para las variables lactosa y ácido láctico presentan un ajuste muy cercano a los datos experimentales pues las dinámicas y tendencias se recuperar de manera muy razonable. En lo que a las distribuciones posteriores se refiere (Figura 5.12), siguen habiendo grandes regiones de probabilidad que generan intervalos de confianza muy amplios, además de la persistencia de la multimodalidad. En general los parámetros respectivos para las variables lactosa y ácido láctico poseen distribuciones donde las modas de las distribuciones respectivas están más definidas, caso que no ocurre para los parámetros de la variable biomasa.

Cuadro 5.5: Parámetros estimados por la moda de la distribución posterior (Figura 5.12) para el modelo (5.5) en un subintervalo de tiempo $[0, 17]$.

Parámetro	$umax$	ks	$Xmax$	a	b	Yps	ms
Estimación	0.8583	15.3519	1.4062	7.9881	0.6611	8.4188	2.8039

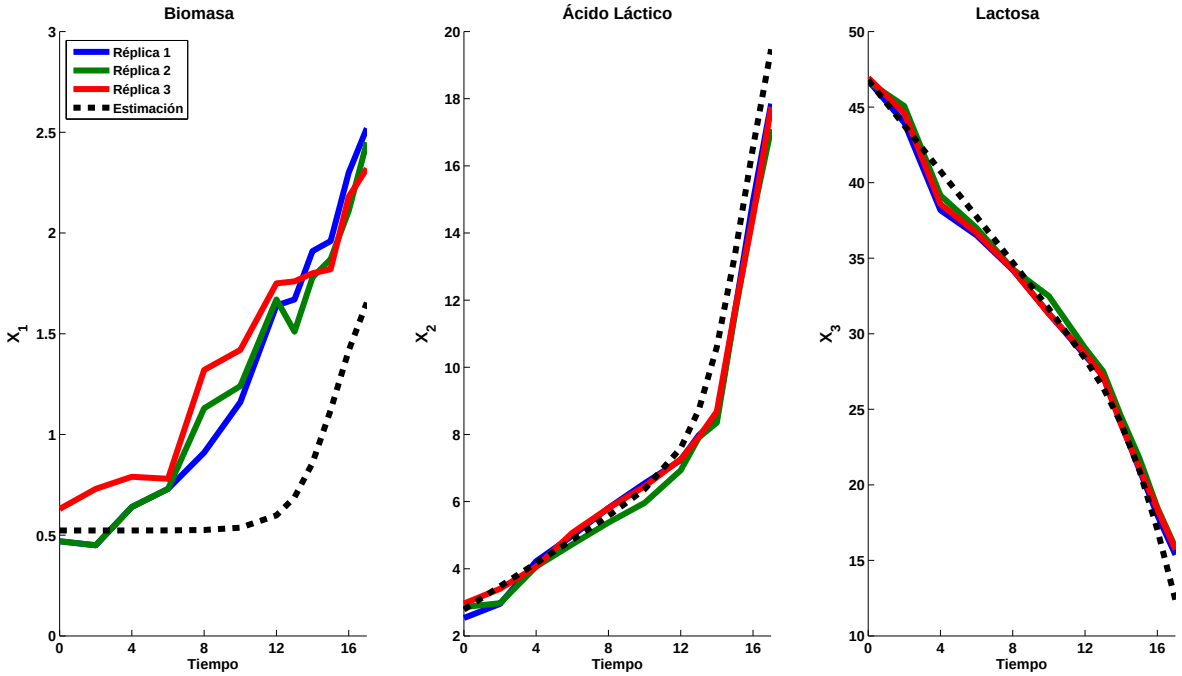

Figura 5.11: Comparativa entre los datos experimentales (líneas de color) y las soluciones estimadas (líneas negras punteadas) para el modelo (5.5) en un subintervalo de tiempo $[0, 17]$.

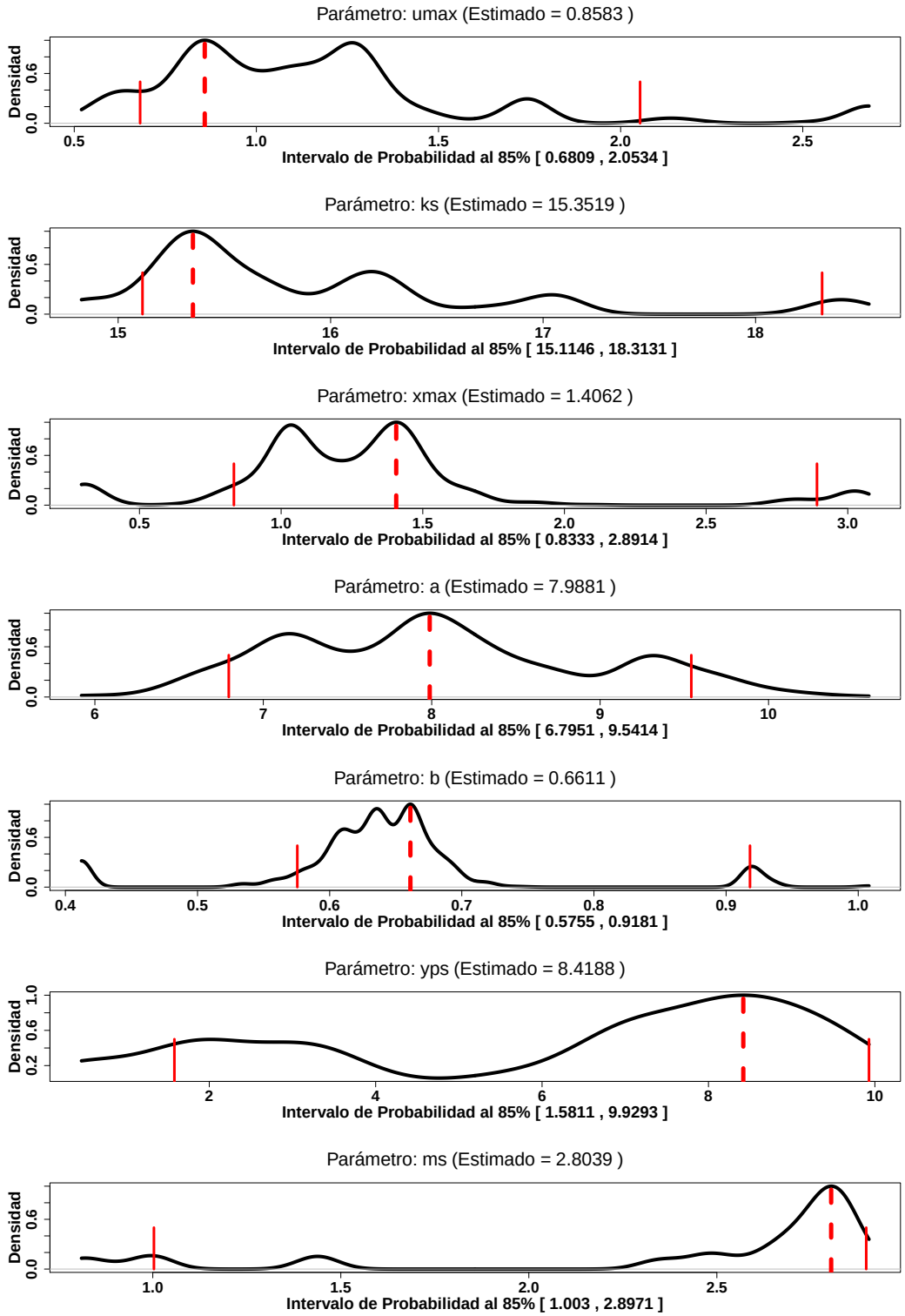

Figura 5.12: Por filas, las distribuciones posteriores de los parámetros del modelo (5.5) en un subintervalo de tiempo $[0, 17]$. La línea roja cortada corresponde al valor estimado por la moda, las líneas rojas sólidas corresponden a los extremos del intervalo de probabilidad al 85 %.

5.5. Modelo exponencial

Un nuevo modelo fue propuesto el cual se denominará modelo exponencial, expresión (5.7). El modelo es similar al logístico salvo la primera ecuación que corresponde al término exponencial en la derivada de la biomasa. La estimación se llevó a cabo considerando los datos experimentales en el intervalo $(0, 17]$ pues durante la estimación se presentan problemas de valores asintóticos en cero para ciertos valores de los parámetros. Los parámetros iniciales se muestran en la expresión (5.8).

$$\begin{aligned} \frac{dX_1}{dt} &= \exp\left(\left(\frac{t}{v_2}\right)^{v_1}\right) \cdot \frac{v_1}{v_2} \cdot \left(\frac{t}{v_2}\right)^{v_1-1} \\ \frac{dX_2}{dt} &= a \cdot \frac{dX_1}{dt} + b \cdot X_1 \\ \frac{dX_3}{dt} &= -\frac{1}{Yps} \cdot \frac{dX_2}{dt} - ms \cdot X_1 \end{aligned} \tag{5.7}$$

$$\boldsymbol{\theta}_0 = \begin{bmatrix} v_1 \\ v_2 \\ a \\ b \\ Yps \\ ms \end{bmatrix} = \begin{bmatrix} 1.13 \\ 16 \\ 8.781 \\ 0.671 \\ 0.581 \\ 0.865 \end{bmatrix} \tag{5.8}$$

Estimación y Solución

El Cuadro 5.6 muestra los parámetros estimados para el modelo (5.7). Las soluciones estimadas se muestran en la Figura 5.13 y se observa que para la variable biomasa se subestiman los datos. Para las variables restantes se tiene un buen ajuste durante el corto y mediano plazo de la dinámica para posteriormente perder la dinámica. Respecto a las distribuciones posteriores (Figura 5.14) se observa que las mismas se encuentran relativamente concentradas alrededor de la moda de la distribución y aunque parecen estar algo dispersas en su forma poseen intervalos de probabilidad más reducidos respecto a estimaciones anteriores.

Cuadro 5.6: Parámetros estimados por la moda de la distribución posterior (Figura 5.14) para el modelo (5.7).

Parámetro	v_1	v_2	a	b	Yps	ms
Estimación	5.4313	20.2196	8.1322	0.5429	16.0582	2.6983

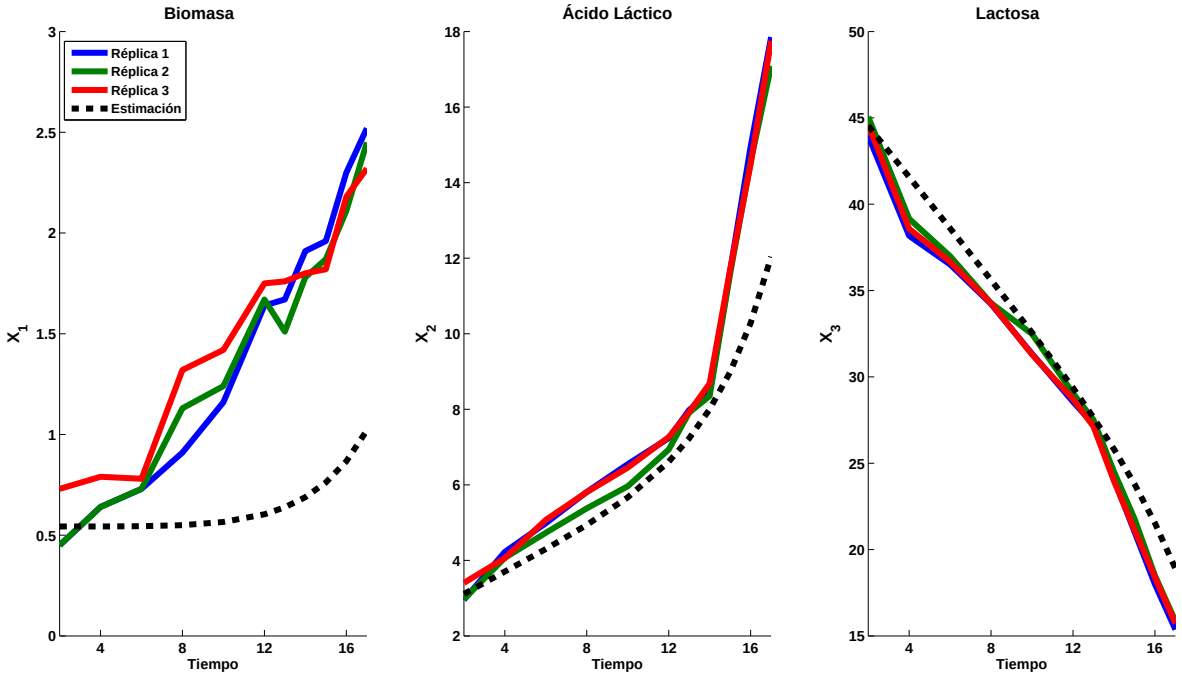

Figura 5.13: Comparativa entre los datos experimentales (líneas de color) y las soluciones estimadas (líneas negras punteadas) para el modelo (5.7).

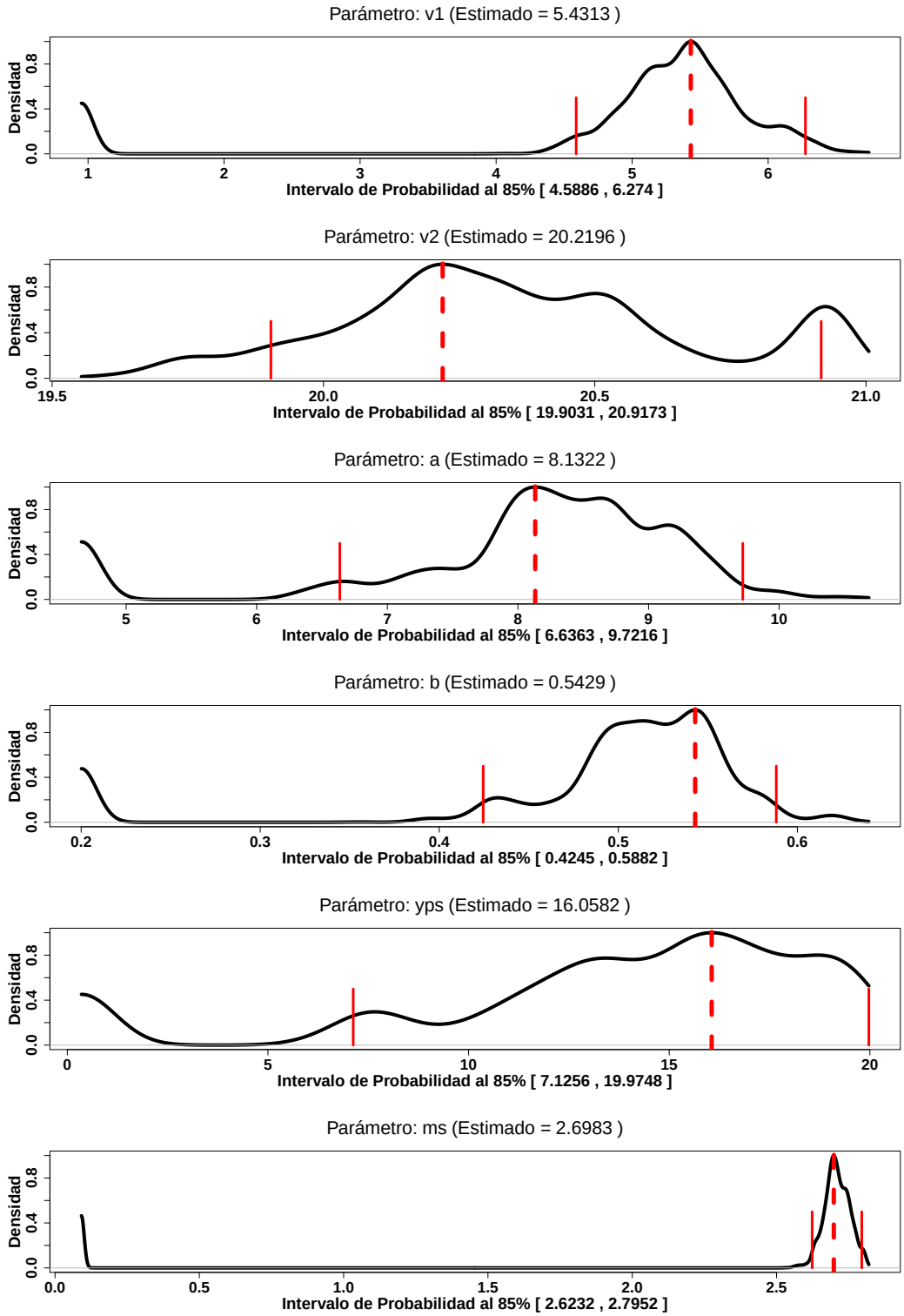

Figura 5.14: Por filas, las distribuciones posteriores de los parámetros del modelo (5.7). La línea roja cortada corresponde al valor estimado por la moda, las líneas rojas sólidas corresponden a los extremos del intervalo de probabilidad al 85 %.

5.6. Modelo logístico antes de la aceleración del ácido láctico

Se finalizó el proceso de estimación y de modelado al considerar nuevamente el modelo logístico (5.5) ahora en un intervalo de tiempo más reducido $[0, 14]$, que corresponde a la etapa de la dinámica antes de la aceleración del ácido láctico, X_2. Se optó por conservar los parámetros de la variable biomasa (X_1, k, tc y $Xmax$) fijos tras haberlos estimado y lograr un buen ajuste como se muestra en la expresión (5.9). Así que únicamente se estimarán los parámetros a, b, Yps y ms.

$$\boldsymbol{\theta}_0 = \begin{bmatrix} k \\ tc \\ Xmax \\ a \\ b \\ Yps \\ ms \end{bmatrix} = \begin{bmatrix} 0.1642 \\ 15.3006 \\ 3.9282 \\ 7.31 \\ 0.659 \\ 0.57 \\ 0.865 \end{bmatrix} \tag{5.9}$$

Estimación y Solución

Las soluciones estimadas (Figura 5.15) para las variables biomasa y ácido láctico presentan un buen ajuste a los datos a lo largo de toda la dinámica, sin embargo para el caso de la variable lactosa la tendencia se recupera pero no así la dinámica en el mediano plazo. Es importante notar que las distribuciones posteriores de los parámetros (Figura 5.16) resultan ser unimodales, con intervalos de probabilidad reducidos y por por tanto más informativos que el de las distribuciones anteriormente estimadas.

Cuadro 5.7: Parámetros estimados por la moda de la distribución posterior (Figura 5.16) para el modelo (5.5) en el intervalo de tiempo $[0, 14]$.

Parámetro	a	b	Yps	ms
Estimado	3.9448	0.0131	0.2705	0.011

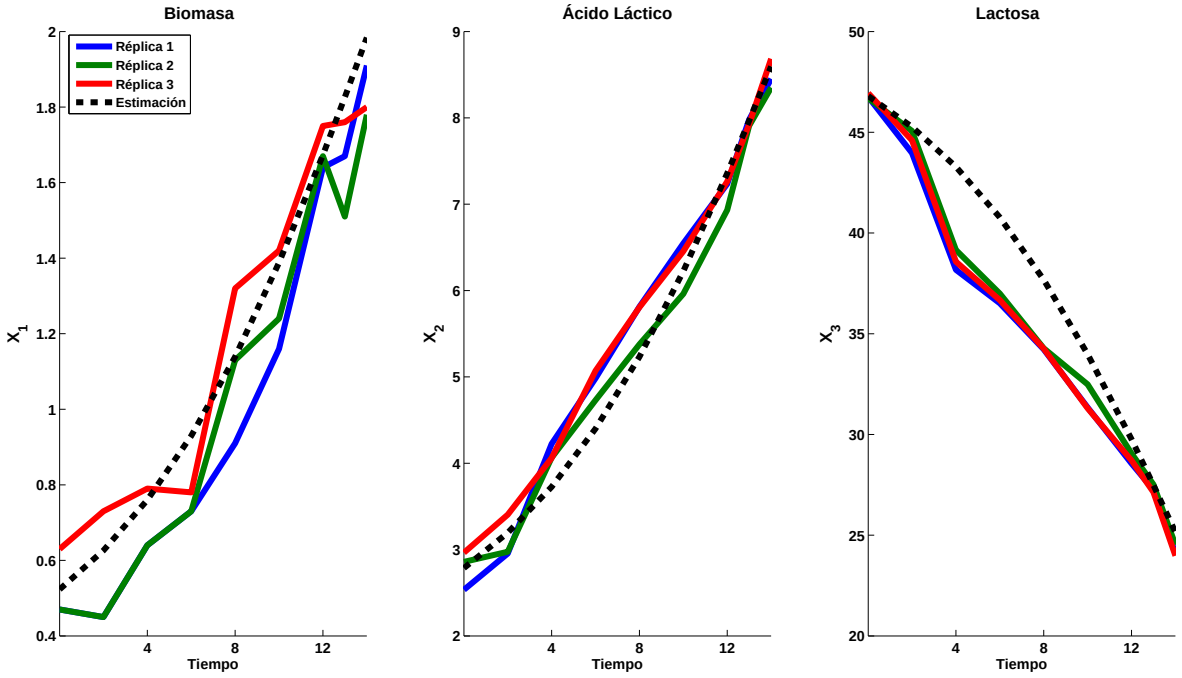

Figura 5.15: Comparativa entre los datos experimentales (líneas de color) y las soluciones estimadas (líneas negras punteadas) para el modelo (5.5) en el intervalo de tiempo $[0, 14]$.

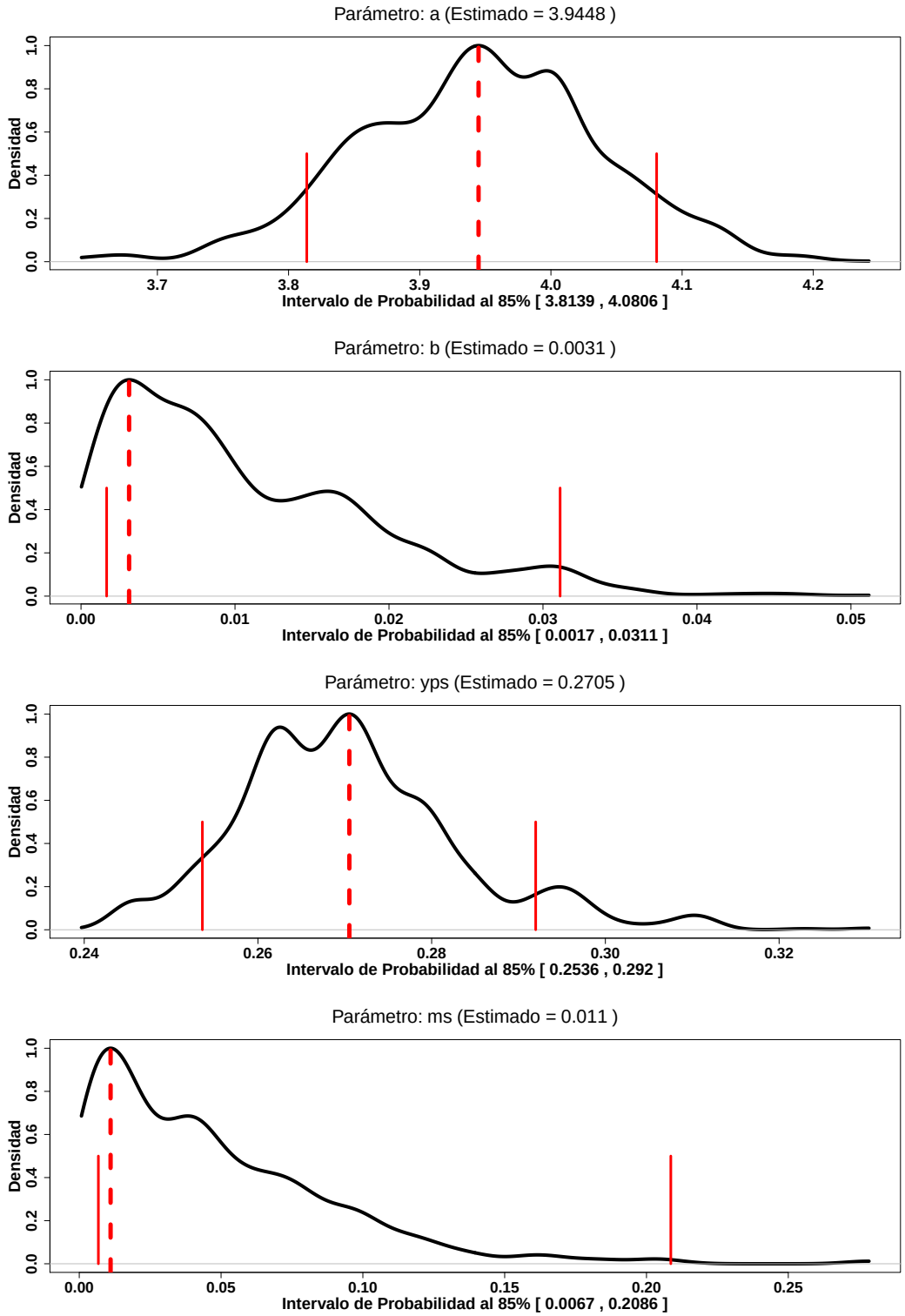

Figura 5.16: Por filas, las distribuciones posteriores de los parámetros del modelo (5.5) en el intervalo de tiempo $[0, 14]$. La línea roja cortada corresponde al valor estimado por la moda, las líneas rojas sólidas corresponden a los extremos del intervalo de probabilidad al 85 %.

5.7. Resultados

El proceso anterior de estimación y modelado reflejó, en primer lugar, la carencia de ajuste para la variable ácido láctico, X_2, seguido de la variable X_1, biomasa. Tales problemas de ajuste se intentaron solucionar al modificar el modelo y reducir el intervalo de datos para considerar regiones de la dinámica menos complejas. De Chávez y Castaño (2014), se justifica que tal reducción de datos experimentales y por tanto en la complejidad en las reacciones, se ve reflejada por medio de las distribuciones posteriores al ser menos dispersas, y así mismo, en soluciones con un mejor ajuste a los datos.

Se concluye que los modelos propuestos aún carecen de la suficiente flexibilidad para tratar con las dinámicas expuestas aún cuando trabajan razonablemente en regiones con dinámicas simples, donde seguirían siendo modelos muy sobrados para ese tipo de estimación. Se recomendaría proponer un nuevo tipo de modelo.

6. Aplicación: Hidrólisis

La aplicación de la sección 5 mostró los problemas que se presentan durante el modelado de un fenómeno, sin embargo, la presente aplicación (como un ejemplo adicional) permite destacar los efectos del modelado inadecuado a partir de supuestos teóricos idealizados. Se presenta la estimación de dos modelos de un mecanismo de hidrólisis de proteína. En Martínez *et al*. (2011) se pueden obtener los detalles del experimento, sin embargo, para nuestro caso únicamente se realizó la estimación de los modelos correspondientes a datos de grado de hidrólisis con concentración de enzima inicial E_0 de 6.36 AU (Anson units)/L) y de concentración de sustrato inicial S_0 de 18.73 g/L, para datos sujetos a un pre-tratamiento.

Bajo algunos supuestos paradigmáticos (Martínez *et al*., 2011), Gónzalez-Tello *et al*. (1994) propone el modelo (6.1), donde ϵ es un parámetro dependiente de la presencia de un sustrato inhibidor y k_d un parámetro referente a la desnaturalización de la enzima.

$$\frac{dh}{dt} = k_d \cdot \left(\frac{E_0}{S_0} - \epsilon\right) \cdot \exp\left(-\frac{k_d}{k_2} \cdot h\right) \tag{6.1}$$

En Martínez *et al*. (2011) se menciona que no se pudo lograr una estimación exitosa de los datos a partir del modelo (6.1) puesto que se presentaron problemas de convergencia. Dada su simplicidad se logró obtener una solución analítica del mismo, resultando en la expresión siguiente:

$$h\left(t\right) = \frac{k_2}{k_d} \cdot ln\left[t \cdot \frac{k_d^2}{k_2} \cdot \left(\frac{E_0}{S_0} - e\right) + 1\right]$$

Para que la expresión anterior esté bien definida se deben considerar la presencia de asíntotas en la función logaritmo natural (ln). Es decir, se deben evitar aquellos valores que cumplan la expresión siguiente:

$$t \cdot \frac{k_d^2}{k_2} \cdot \left(\frac{E_0}{S_0} - e\right) + 1 \leq 0$$

La única restricción del modelo (6.1) es que el parámetro e sea menor a E_0/S_0, que en este caso sería de 0.3396.

A manera de contraste y dadas las dificultades para lograr una estimación exitosa debido a problemas de convergencia, en Martinez *et al*. (2011) se propuso un nuevo modelo definido por la expresión (6.2).

$$\frac{dh}{dt} = a \cdot \exp\left(-b \cdot h\right) \tag{6.2}$$

6.1. Estimación

Para el proceso de estimación se utilizó el método population-based MCMC. Se otorgaron a prioris uniformes en el intervalo $[0, 100]$ para los parámetros a, b, k_d y k_2; mientras que una a priori uniforme $[0, 0.339]$ para el parámetro e, por las razones ya mencionadas.

Las Figuras 6.1 y 6.2 muestran las distribuciones posteriores para los parámetros estimados de los modelos (6.1) y (6.2), respectivamente. Se observa que para el modelo (6.1) las distribuciones posteriores son claramente bimodales para los parámetros k_d y k_2 (Figura 6.1, primer y segundo panel). Los intervalos de probabilidad resultan ser muy amplios así como también ocurre para los intervalos de probabilidad del parámetro e, que además, presenta una alta densidad de valores alrededor del valor crítico 0.3396, (Figura 6.1, tercer panel). Una situación contraria se presenta para el modelo (6.2), donde las distribuciones posteriores son claramente unimodales con intervalos de probabilidad muy reducidos y simétricos, Figura 6.2.

La Figura 6.3 muestra las soluciones estimadas (línea negra punteada) para los datos experimentales (líneas sólidas). Para el caso del modelo (6.1), primer panel, se observa que la solución subestima completamente los datos experimentales desde el inicio de la dinámica; situación que no ocurre para el modelo (6.2), segundo panel, donde se obtiene una estimación lo bastante adecuada para los datos experimentales en sus dos réplicas.

6.2. Resultados

Los datos experimentales, como se puede observar (Figura 6.3) poseen poca variabilidad en su dinámica (dinámicas muy suaves), es por ello que se pueda concluir que lo reflejado en las distribuciones posteriores de los parámetros del modelo, en la forma de la distribución y en la amplitud de sus intervalos de probabilidad (6.1), no es más que un problema de especificación de modelo, más que el efecto del ruido muestral mismo.

Si bien muchos de los modelos utilizados en la práctica son producto del desarrollo teórico (manifestaciones de una teoría); esta teoría se basa generalmente en condiciones experimentales inverosímiles (o raramente posibles de alcanzar) para los fenómenos. La alternativa, el modelaje empírico de un fenómeno, si bien puede perder las bases teóricas para la interpretación, es una alternativa sólida en las aplicaciones, pero que requiere un esfuerzo adicional para construir e interpretar a los parámetros que llegara a involucrar.

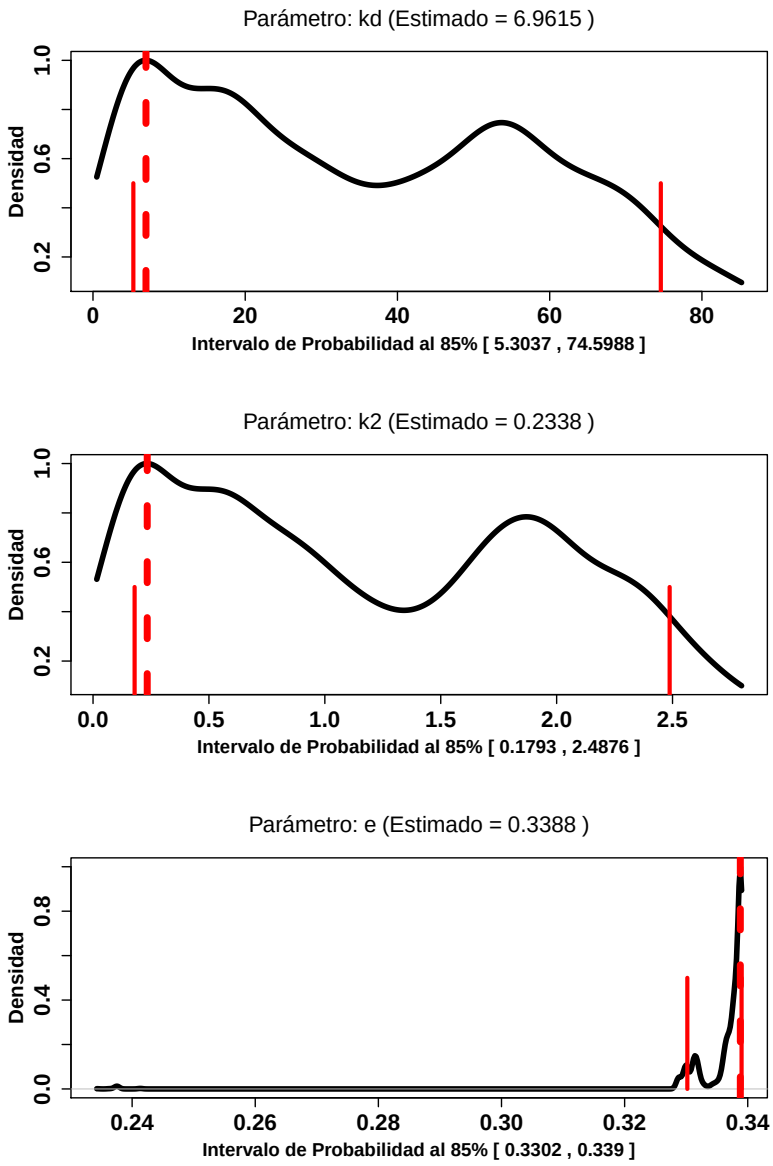

Figura 6.1: Por filas, las distribuciones posteriores de los parámetros del modelo (6.1). La línea roja cortada corresponde al valor estimado por la moda, las líneas rojas sólidas corresponden a los extremos del intervalo de probabilidad al 85 %.

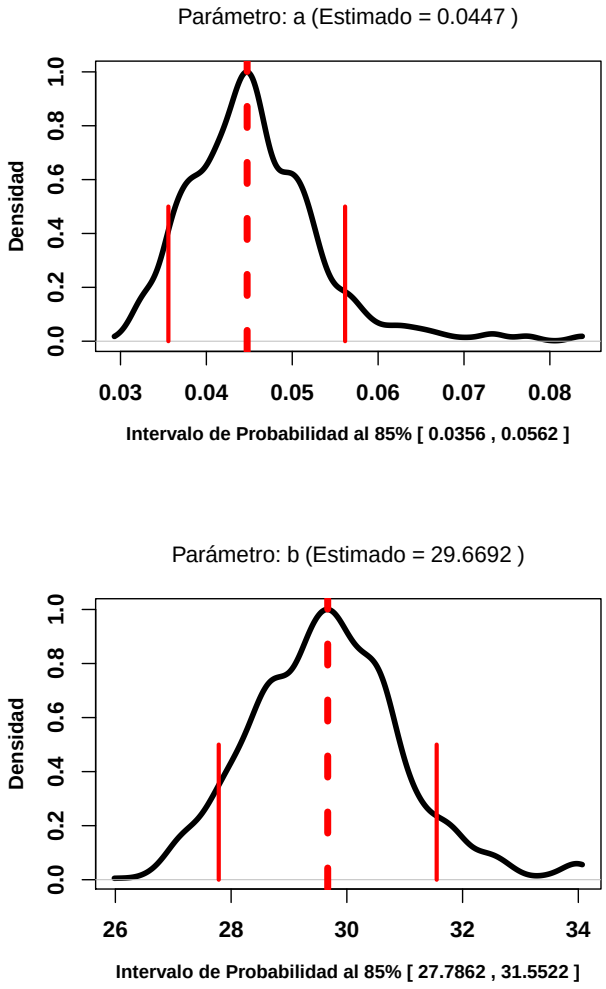

Figura 6.2: Por filas, las distribuciones posteriores de los parámetros del modelo (6.2). La línea roja cortada corresponde al valor estimado por la moda, las líneas rojas sólidas corresponden a los extremos del intervalo de probabilidad al 85 %.

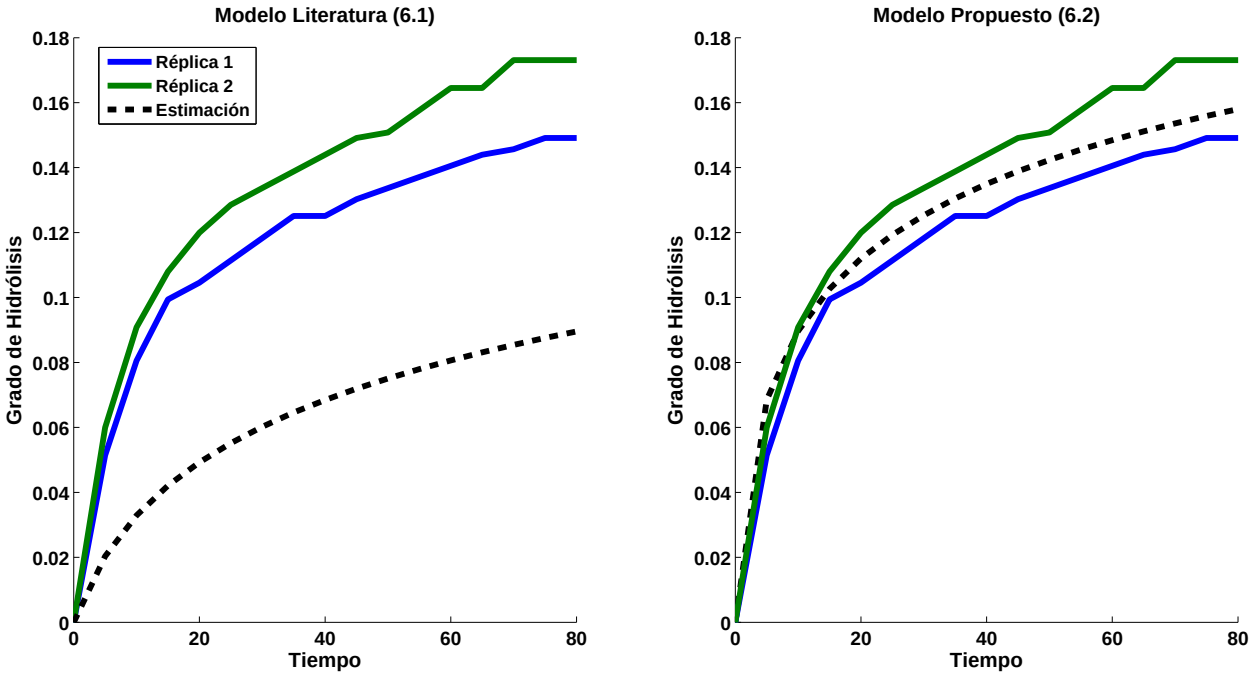

Figura 6.3: Comparativa entre los datos experimentales (líneas de color) y las soluciones estimadas (líneas negras punteadas) para los modelos (6.1) y (6.2), primer y segundo cuadro, respectivamente.

7. Conclusiones

Comparativa de Métodos de Estimación

Los métodos de estimación de parámetros Bayesianos poseen la capacidad de capturar la incertidumbre proveniente del contexto del fenómeno a través de las distribuciones posteriores de sus parámetros. Dentro de estos métodos se destacan aquellos basados en muestreo, como lo son los métodos Metropolis-Hastings y population-based MCMC aquí utilizados. Tales algoritmos fueron puestos a prueba bajo distintos escenarios que comprenden la estimación de parámetros en modelos con diversa complejidad así como el manejo de distintos niveles de ruido muestral.

Durante este proceso de estimación se observaron algunas de las fortalezas y deficiencias de estos algoritmos. Por un lado, el método Metropolis-Hastings presentó problemas de convergencia de la distribución posterior hacia un conjunto de valores óptimos de los parámetros ante sistemas dinámicos oscilatorios. El algoritmo corre un gran riesgo de quedar estancado en modas subóptimas de la distribución posterior acarreando el ya mencionado problema de trampa local (Faming *et al*., 2010). Tal situación se ve evitada por el uso del algoritmo population-based MCMC, permitiendo discriminar entre modas subóptimas, logrando así proveer de mejores resultados en la estimación. Ante tales situaciones son evidentes ahora los escenarios sobre los cuales los métodos se desempeñas de mejor manera, por un lado, ante modelos simples (estables en el tiempo o con dinámicas no oscilatorias) y con suficiente información a priori del fenómeno, el algoritmo Metropolis-Hastings resulta ser suficiente para obtener buenos resultados en la estimación, mientras que para modelos complejos, y con insuficiente información, como lo fue el sistema presa depredador en su vertiente oscilatoria, el algoritmo population-based MCMC resulta ser el más adecuado en el sentido de arrojar estimaciones de los parámetros que permiten recuperar la dinámica de los datos experimentales, lidiando con las fuentes de incertidumbre presentes. La comparativa realizada entre ambos métodos, determinó la elección del algoritmo population-based MCMC como el adecuado para realizar el proceso de estimación de parámetros a lo largo de la investigación debido a que se trabajaría con modelos caóticos (sistema de Lorenz) y escenarios donde la información del fenómeno es escasa o no la teoría no se desempeña adecuadamente (aplicación del biorreactor e hidrólisis).

Sistemas Caóticos

Tras haber contrastado los métodos de estimación en base a su desempeño ante sistemas de diversa complejidad surge la idea de utilizar este tipo de estimación Bayesiana en sistemas complejos de tipo caótico, como lo es el sistema de Lorenz, y observar el desempeño del método Bayesiano ante estos fenómenos. El sistema de Lorenz posee dos vertientes en su dinámica dependiente de los valores de sus parámetros, un estado no caótico (pero oscilatorio) y un estado caótico.

Los resultados del proceso de estimación sobre el sistema de Lorenz se ven reflejados en la forma de las distribuciones posteriores de los parámetros, por un lado, en la estimación sobre el sistema en su estado no caótico se presentaron distribuciones posteriores unimodales que a grandes rasgos únicamente variaron en su dispersión al cambiar el valor de la desviación estándar en el ruido gaussiano añadido en los datos, resultando en estimadores que generan soluciones bastante cercanas a los datos experimentales. Esta situación contrasta con aquellas estimaciones realizadas en el sistema en su estado caótico, donde las distribuciones posteriores fueron claramente multimodales; situación que dificultó la obtención de una estimación para los parámetros dando como resultado estimaciones de las soluciones totalmente alejadas de los datos experimentales en el corto plazo, independientemente del ruido muestral. Esta situación es únicamente provocada por el carácter caótico del sistema, debido a que se tiene un modelo bien especificado con datos únicamente sujetos a ruido muestral.

Tal carencia de ajuste a los datos no es el único resultado producto de la estimación en un sistema caótico, se observó además que independientemente del ruido muestral y de la complejidad del modelo, ya sea el sistema en su estado caótico o no caótico, el método Bayesiano arroja un estimador de los parámetros que logra recuperar la tendencia general (a largo plazo) del atractor del sistema. Aterrizando estos resultados en el contexto meteorológico sobre el cual está formulado el sistema de Lorenz, se confirma que ante sistemas caóticos una estimación a corto plazo de un fenómeno como lo es el clima es imposible, no así una estimación del comportamiento visualizado en una perspectiva temporal amplia.

Aplicaciones

Finalmente, respecto al par de aplicaciones realizadas se logró enfrentar el método Bayesiano ante los problemas propios de la práctica experimental como lo son la falta de información ante un fenómeno y los problemas de especificación de un modelo. En la primer aplicación que corresponde a los fenómeno ocurridos en un biorreactor (sección 5) se presentó un proceso que implicaba proponer y modificar un modelo con el fin de capturar las dinámicas inmiscuidas en el fenómeno. De tal proceso se derivó que los modelos propuestos a pesar de su variada complejidad y que en algunos casos se presentan un ajuste razonable, carecen aún de suficiente flexibilidad por lo que se refleja un problema de carencia de ajuste directamente en las distribuciones posteriores de los parámetros, pues en su mayoría fueron multimodales, por lo que queda abierta la modificación o proposición de un modelo adicional.

En este mismo sentido, en la segunda aplicación (sección 6) que implica un proceso de hidrólisis, se contrastaron dos modelos, uno teórico contra uno empírico. Tal contraste deja claro que no siempre los modelos teóricos se desempeñan adecuadamente en el campo, por lo que se debe tener precaución al utilizarlos y optar por modelos empíricos.

Ambas aplicaciones dejan claro que los métodos de estimación no pueden lidiar con problemas de falta de especificación, sólo arrojan pistas sobre la presencia del mismo, sin embargo tal situación no es tomada en cuenta y se opta por considerar únicamente los valores producto de la estimación sin ver lo que existe de fondo.

Bibliografía

BioBayes: A software package for Bayesian inference in systems biology. Vladislav Vyshemirsky and Mark Girolami. Department of Computing Science, University of Glasgow, G12 8QQ, UK, URL http://www.dcs.gla.ac.uk/BioBayes/.

Blanchard, P., Devaney, R. L., and Hall G. R. 1999. Ecuaciones diferenciales. International Thomson Editores. México.

Brooks, S. P. 1998. Markov Chain Monte Carlo method and its application. The Statistician. 47 (Part 1): 69-100.

Brunel, N. J-B. 2008. Parameter estimation of ODE's via nonparametric estimators. Electron. J. Stat. 2: 1242-1267.

Campbell, D. and Steele, R. J. 2012. Smooth functional tempering for nonlinear differential equation models. Stat. Comput. 22:429–443.

Chávez, C. B. and Castaño, T. E. 2014. Effect of sample noise on the parameter estimation of complex dynamic systems. Engineering applications vol 1. IEEE. (Accepted).

Chis, O-T. Banga, J. R. and Balsa-Canto E. 2011. Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods. PLoS ONE. 6(11): e27755.

Chou, I-C. and Voit, E. O. 2009. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math. Biosci. 19 (Issue 2): 57-83.

Congdon, P. 2006. Bayesian Statistical Modelling (2nd ed.). John Wiley & Sons, Ltd. UK.

Dennis G. Z. 1997. Ecuaciones diferenciales con aplicaciones de modelado (6ed). Intemational Thomson Editores. México.

Faming, L., Chuanhai, L., and Raymond, J. C. 2010. Advanced Markov Chain Monte Carlo methods : learning from past samples. John Wiley and Sons Ltd. UK.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2004. Bayesian Data Analysis (2nd ed.). Chapman & Hall/CRC. D.C.

Geyer, C. J. 1992. Practical Markov Chain Monte Carlo. Stat. Sci. 7 (4): 457-472.

Girolami, M. 2008. Bayesian inference for differential equations. Theor. Comput. Sci. 408 (Issue 1): 4-16.

Gugushvil, S. and Klaassen, C. A. J. 2012. $\sqrt{n}$-consistent parameter estimation for systems of ordinary differential equations: bypassing numerical integration via smoothing. Bernoulli. 18 (No.3): 1061-1098.

Heinz-Otto, P., Hartmut, J. and Dietmar, S. 2004. Chaos and Fractals: New Frontiers of Science (2ed). Springer. USA.

Hirsch, M. W., Smale, S. and Devaney, R. L. 2004. Differential equations, dynamical systems, and an introduction to chaos (2ed). Elsevier Academic Press. USA.

Jasra, A., Holmes, C. C. and Stephens, D. A. 2005. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. Stat. Sci. 20 (No. 1): 50–67.

Jasra, A., Stephens, D. A. and Holmes, C. C. 2007. On population-based simulation for static inference. Stat. Comput. 17 (3): 263–279.

Madar, J., Abonyi, J., Roubos, H. and Szeifert, F. 2003. Incorporating Prior Knowledge in a Cubic Spline ApproximationsApplication to the Identification of Reaction Kinetic Models. Ind. Eng. Chem. Res. 42: 4043-4049.

Martínez, A. G., Castaño, T. E., Amaya, L. S. L., Regalado, G. C., Martínez, V. C and Ozimek, L. 2011. Modeling of Enzymatic Hydrolysis of Whey Proteins.Food Bioprocess Technol. doi: 10.1007/s11947-011-0624-5.

Moles, C. G., Mendes, P. and Banga, J. R. 2003. Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods. Genome Res. 13: 2467-2474.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ramsay, J. O., Hooker, G., Campbell, D. and Cao, J. 2007. Parameter estimation for diferential equations: A generalized smoothing approach. J. R. Statist. Soc. Series B. 69: (Part 5): 741–796.

Ramsay, J. O. and Silverman, B. W. 1997. Functional Data Analysis. New York: Springer-Verlag.

Schaber, J. and Klipp, E. 2011. Model-based inference of biochemical parameters and dynamic properties of microbial signal transduction networks. Curr. Opin. Biotechnol. 22:109-116.

Scheinerman, E. R. 1996. Invitation to dynamical systems. Prentice-Hall. USA.

Spiegel, M. R. 1983. Ecuaciones diferenciales aplicadas (3ed). Prentice-Hall. México.

Stuart, A. M. 2010. Inverse problems: A Bayesian perspective. Acta Numerica, 19: 451-559.

van Riel N. A.W. 2006. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. Brief. Bioinform. 7 (No. 4): 364-374.

Vanlier, J., Tiemann, C. A., Hilbers, P. A. J. and van Riel, N. A. W. 2013. Parameter uncertainty in biochemical models described by ordinary differential equations. Math. Biosci. 246 (Issue 2): 305-314.

Varah, J. M. 1982. A Computational Approach to Parameter Estimation in Ordinary Differential Equations. University of Waterloo. Computer Science Dept. Ontario, Canada.

Varziri, M. S., McAuley, K. B. and McLellan, P. J. 2008. Approximate Maximum Likelihood Parameter Estimation for Nonlinear Dynamic Models: Application to a Laboratory-Scale Nylon Reactor Model. Ind. Eng. Chem. Res. 47: 7274–7283.

Vyshemirsky, V. and Girolami. M. 2008. BioBayes: A software package for Bayesian inference in systems biology. Bioinformatics. 24 (No. 17): 1933–1934.