

EXPERIMENTOS Y RESULTADOS

En los capítulos anteriores se ha descrito los diferentes algoritmos que se utilizarán para realizar la tarea de *speaker diarisation*, y que en esta sección se emplearán de acuerdo al marco experimental que se describe a continuación.

Inicialmente, las pruebas consistieron en usar los algoritmos presentados para selección de modelo usando datos que fueron generados aleatoriamente a partir de los parámetros de un HMM inicial; para tener una idea general de su desempeño individual.

Para estas primeras pruebas, se simuló una cadena de Márkov oculta con en base en parámetros fijos, con lo que se generó tanto una secuencia de datos observados, como los supuestos datos o variables ocultas que forman la cadena de Márkov. Se utilizó muestreo ancestral para la simulación de estos datos.

Para un caso en específico, se tiene lo siguiente. Realizando la inferencia de parámetros del HMM, se obtienen los siguientes resultados:

El primer algoritmo que se prueba, es el de selección de modelo usando un BIC.

Como ya se comentó, se usará una variante de BIC en donde se incorpora un término de regularización λ para que correspondan en órdenes de magnitud tanto la log-verosimilitud del modelo encontrado como su penalización respectiva.

El problema inmediato que se presenta, es cómo realizar la selección del parámetro de regularización λ que penalice de forma correcta la verosimilitud para los diferentes modelos propuestos. Si λ es demasiado pequeño, entonces la penalización realmente no tendrá efecto y dado el sobreajuste que se presenta al usar modelos más complejos, se preferirán siempre los modelos con más parámetros. Por otro lado, si al escoger λ se

da demasiado peso al término de penalización, entonces siempre se preferirán los modelos más sencillos.

Para encontrar el valor de λ adecuado, se puede entonces formar una superficie con las diferentes curvas de selección BIC de acuerdo a cómo varía λ , e inspeccionar esta superficie para encontrar una región de confianza en la que el valor de λ es el adecuado.

Por otro lado, para la segunda prueba, se procedió a usar bootstrap con la estadística log-likelihood ratio como ya se describió anteriormente en el ??, y haciendo la prueba de hipótesis del modelo de n estados contra el de $n + 1$ estados.

5.1 EXPERIMENTOS

Para los experimentos realizados, se generaron mediante un sintetizador de voz (también conocido como Text-To-Speech o TTS por sus siglas en inglés) que nos permitió tener un mayor control sobre el contenido como tal de las grabaciones, así como sobre los posibles ruidos o interferencias en las secuencias de audio.

Si bien, para probar el desempeño contra otras propuestas del estado del arte se suelen usar otro tipo de bases de datos, éstas suelen no estar disponibles de forma libre, por lo que preferimos generar nosotros un pequeño dataset con el sintetizador de voz.

Usando dos motores para el sintetizador de voz, uno con voces en inglés y otro con voces en español, se generaron 6 secuencias de audio (3 en cada idioma) cuya duración así como el número de interlocutores que participan varía.

Secuencia 1: Edgar Allan Poe

En esta primer secuencia, se tomaron varios poemas del escritor Edgar Allan Poe, y se utilizaron 6 diferentes voces en inglés. La secuencia de audio original es de 12:06m.

Para la etapa de agrupación de los vectores MFCC con k-means++ se usaron 140 centros iniciales.

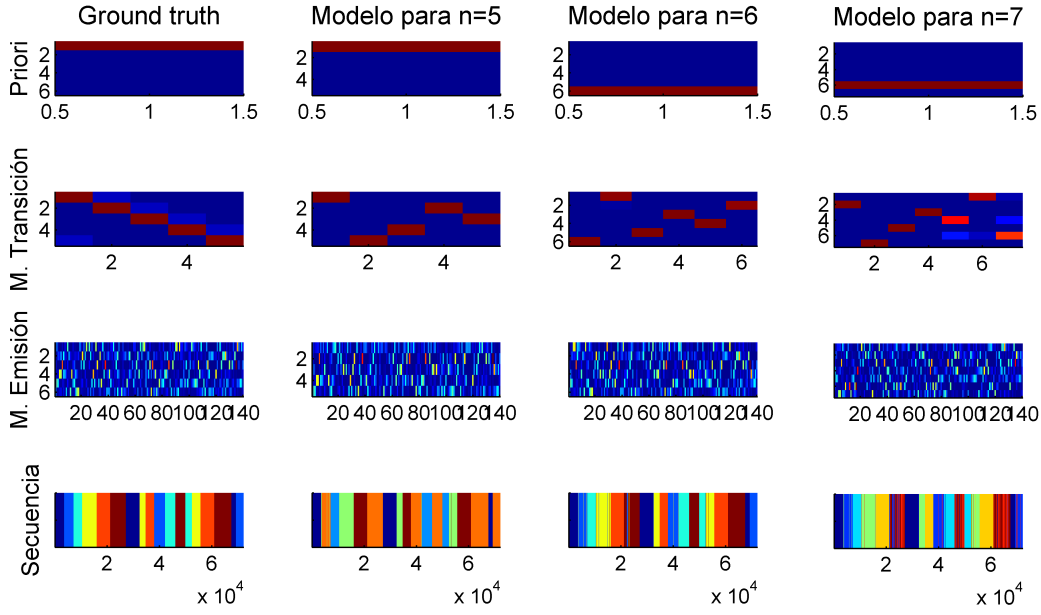


Figura 5.1: *Parámetros encontrados para Secuencia 1.*

En la [Figura 5.21](#) se muestran los parámetros obtenidos para diferentes modelos que se propusieron. Se observa como para el modelo correcto tanto la matriz de emisión así como la secuencia en falso color corresponden con el ground truth.

Utilizando nuestra metodología, primero realizamos una inspección por medio de BIC regularizado para encontrar cuál o cuáles son los modelos más probables.

Se realiza un análisis de sensibilidad primero para determinar cuál es el parámetro λ adecuado que penaliza de forma adecuada la logverosimilitud. Para esto, primero realizamos:

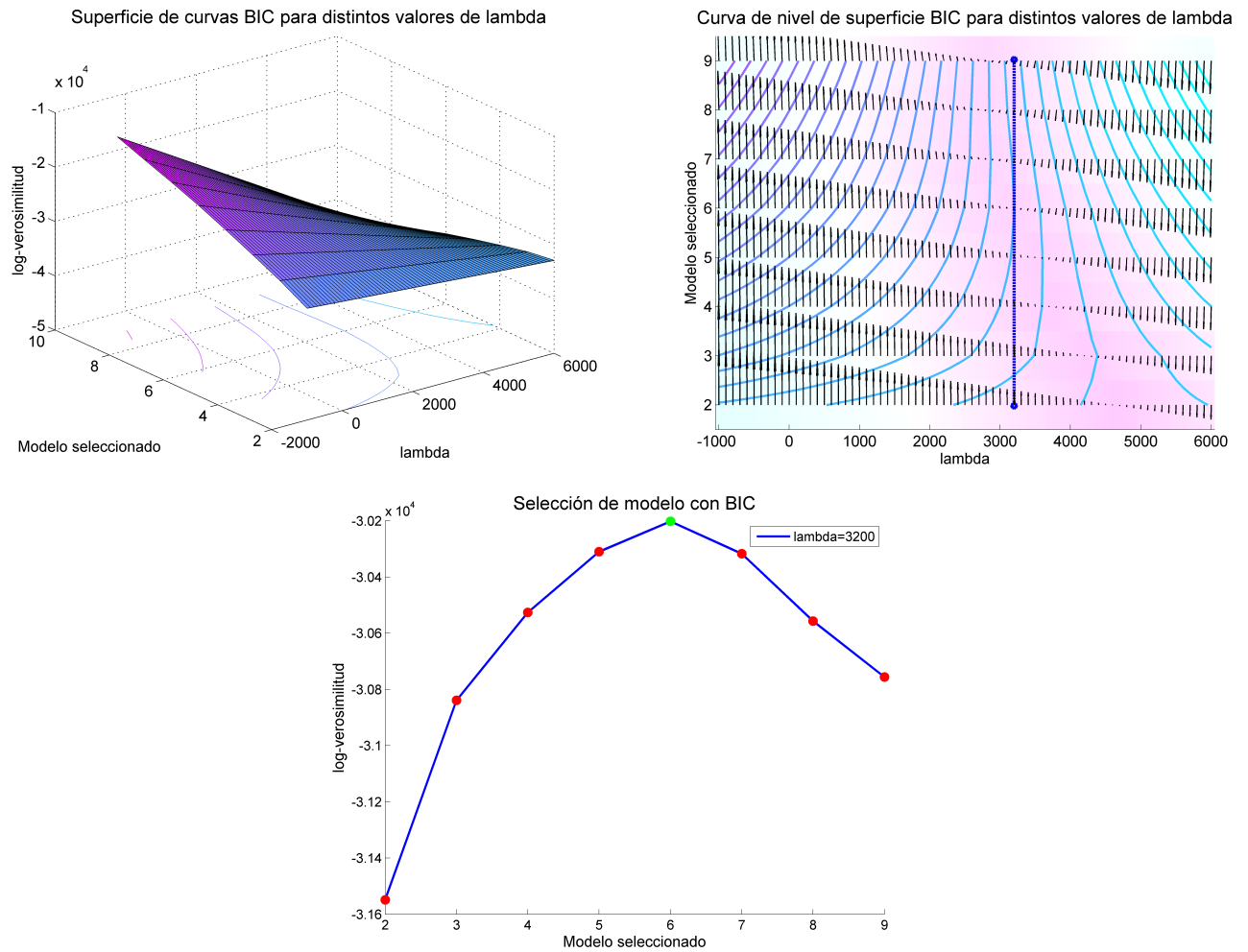


Figura 5.2: Superficie y curva de nivel BIC para Secuencia 1.

En la primera imagen, se muestra la superficie generada variar el valor de λ para evaluar BIC. En la segunda imagen, se muestra las curvas de nivel de la superficie anterior, así como la dirección del gradiente en la misma. En la tercera imagen se muestra el valor seleccionado para λ de acuerdo al análisis de sensibilidad realizado.

Después, planteamos una prueba de hipótesis para ver qué tan acertada fue nuestra selección de modelo. Realizaremos la prueba de hipótesis para los modelos que obtuvieron una mejor valoración por BIC, y de ahí discriminaremos cuál es el que se ajusta mejor, utilizando bootstrap

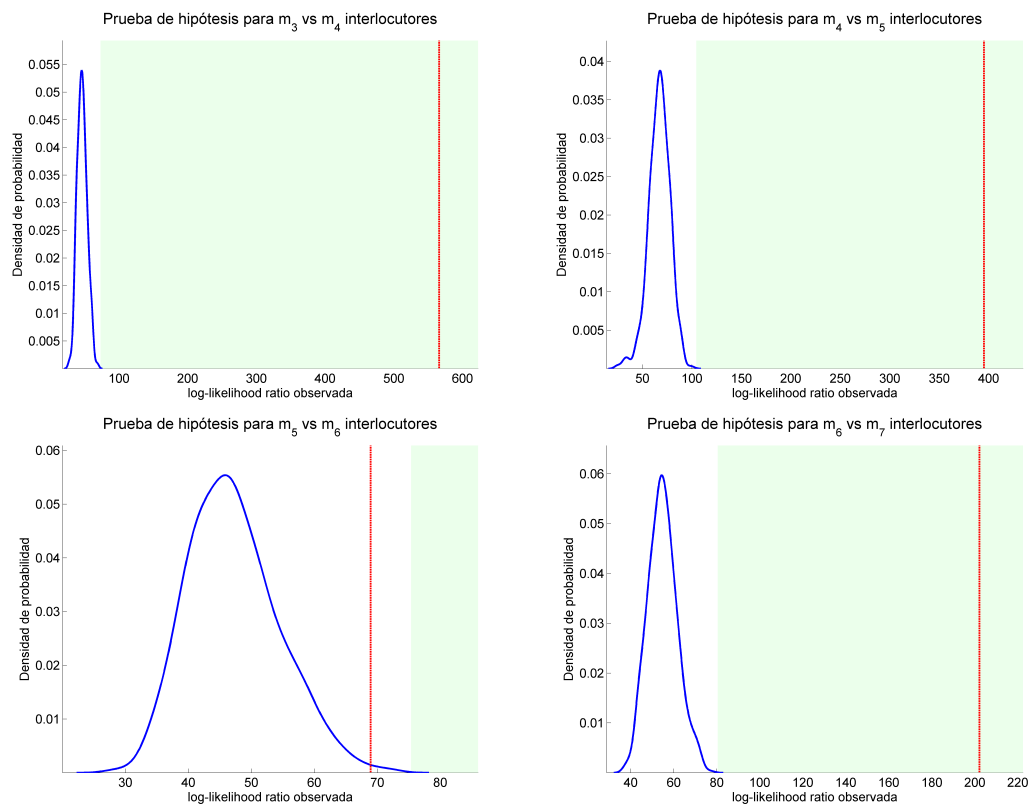


Figura 5.3: *Pruebas de hipótesis con bootstrap para Secuencia 1.*

Por último, ya con el modelo seleccionado, se procede a calcular el error relativo de predicción que se obtuvo, comparando con el ground truth que se dispone para esa secuencia, con lo que se obtiene:

...

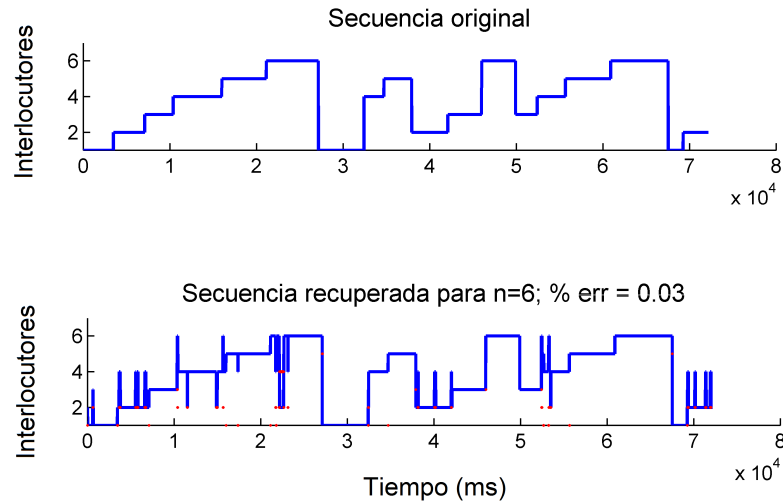


Figura 5.4: *Secuencias encontradas para Prueba 1.*

Más a detalle, en la [Figura 5.24](#) se observa en azul el orden en el que participan los interlocutores de acuerdo a la secuencia recuperada. En rojo se marcan tanto los falsos positivos como los falsos negativos, de acuerdo al ground truth. Hay que notar que cuando el número de estados para un modelo no es el correcto, entonces inminentemente el número de errores en la secuencia obtenida será mayor, pues al menos todas las intervenciones de un hablante no podrán ser emparejadas o serán asignadas a alguien más.

Se observa también que la mayoría de las veces, en la secuencia recuperada se encuentran algunos brincos entre personas, pero en esencia la estructura y el orden en que hablan los interlocutores es el correcto.

Secuencia 2: Gabriel García Márquez

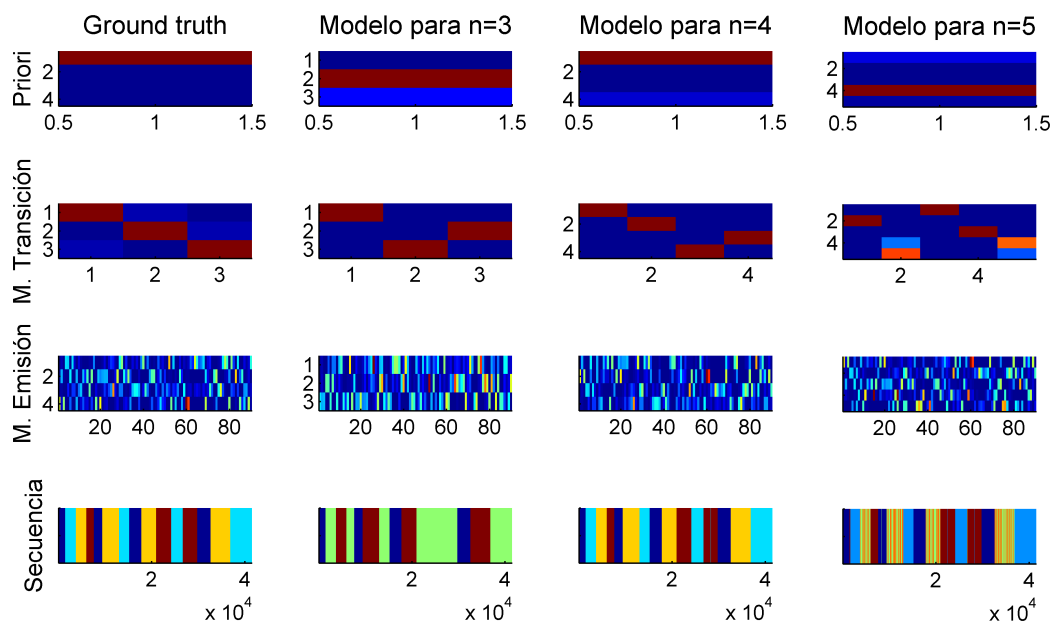


Figura 5.5: *Parámetros encontrados para Secuencia 2.*

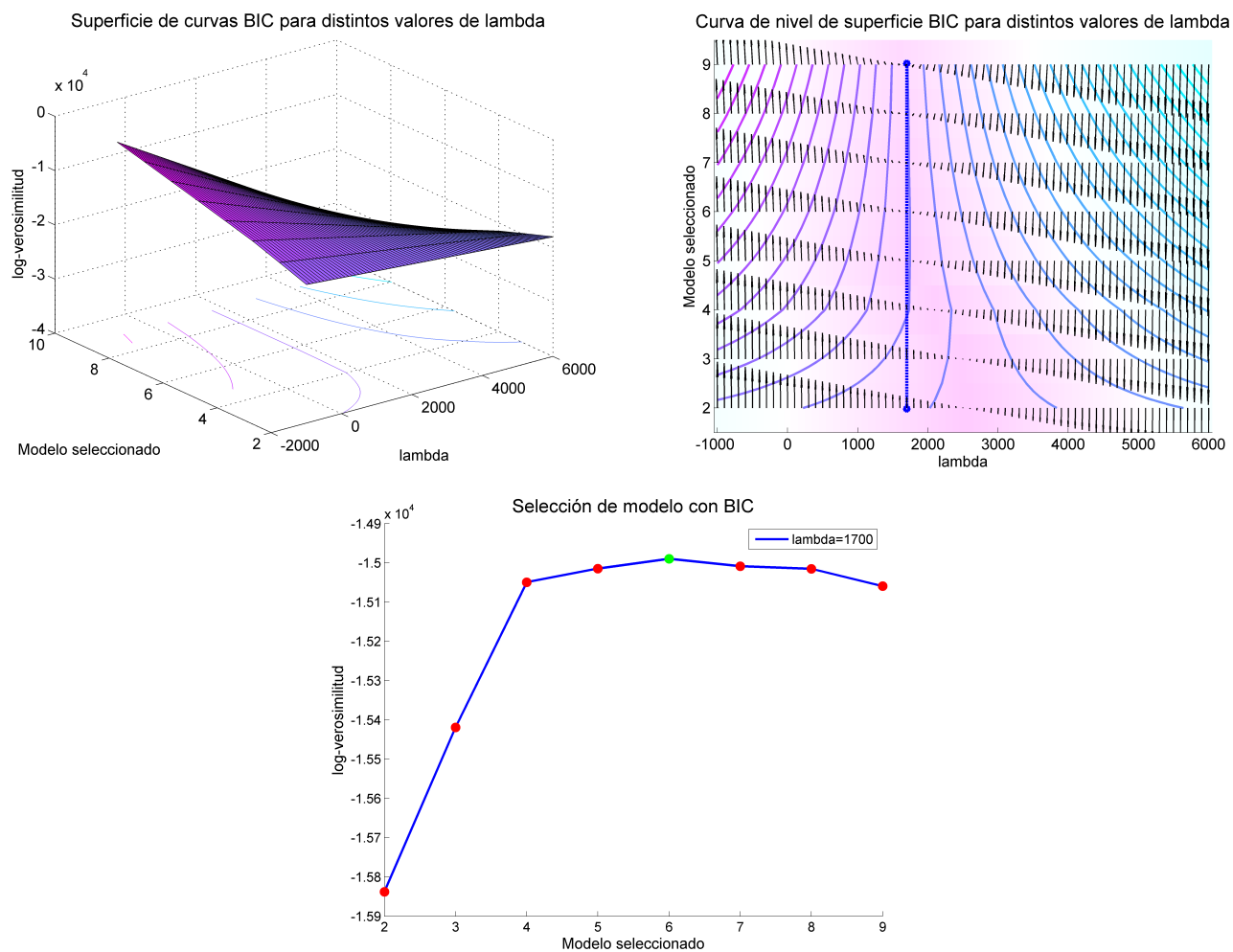


Figura 5.6: Superficie y curva de nivel BIC para Secuencia 2.

Descripción.

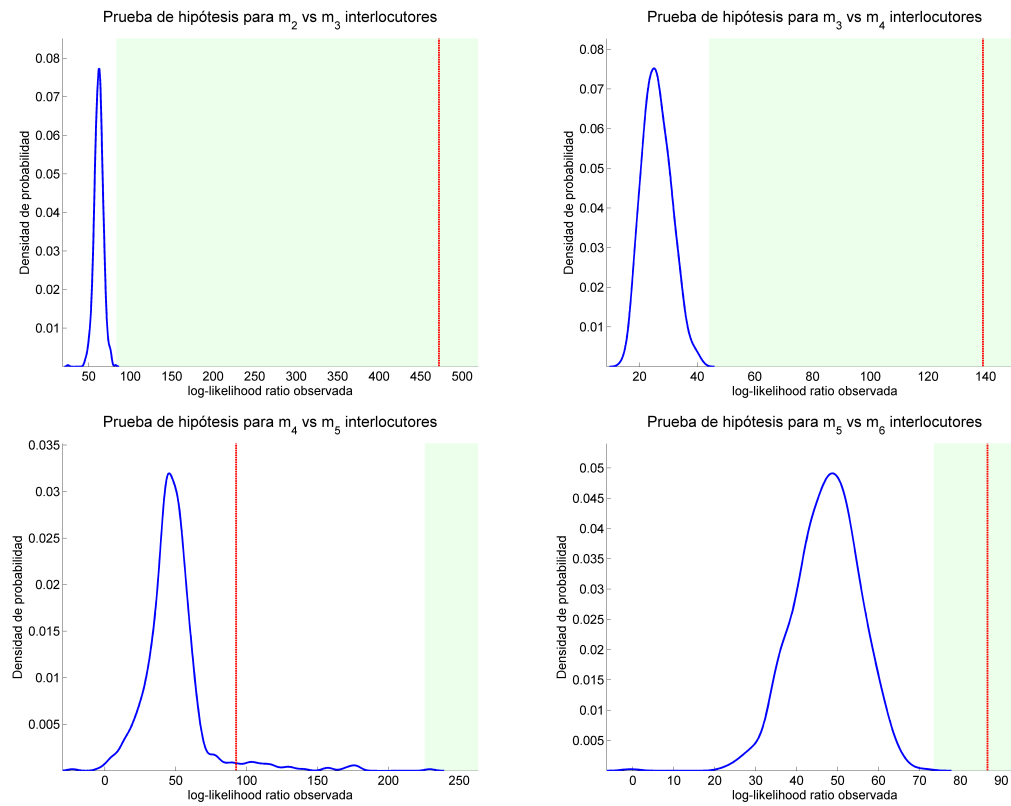


Figura 5.7: Pruebas de hipótesis con bootstrap para Secuencia 2.

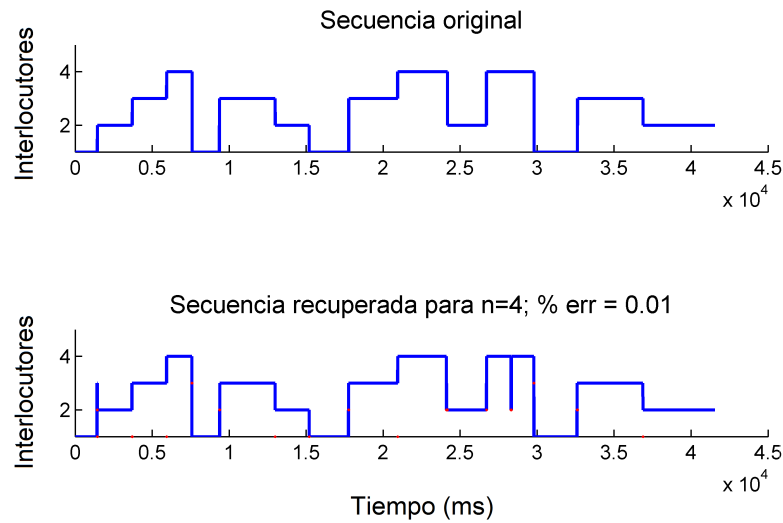
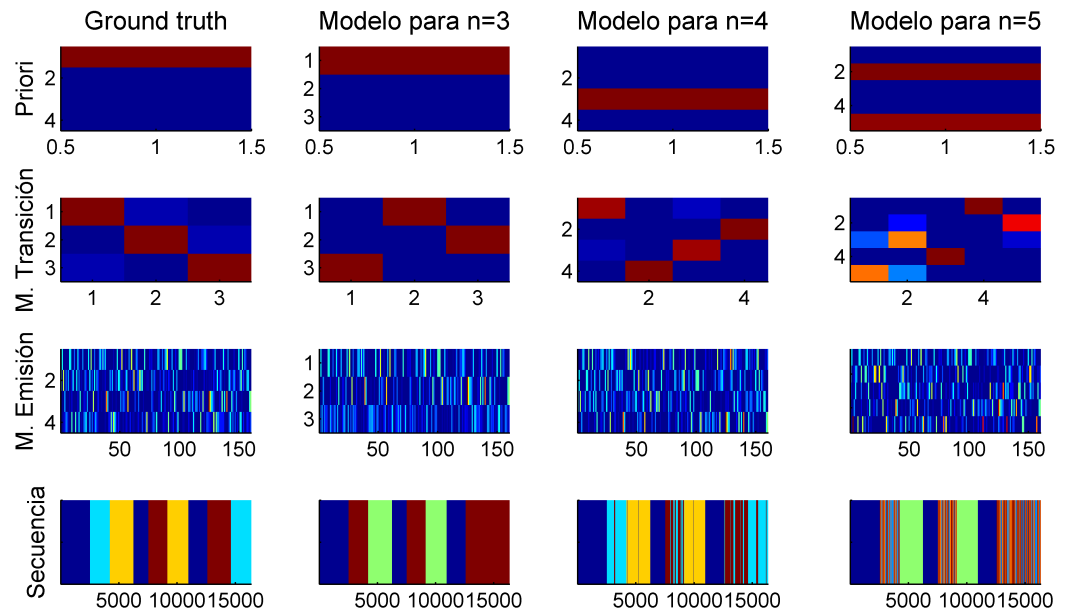


Figura 5.8: Secuencias encontradas para Prueba 2.

Secuencia 3: William Shakespeare**Figura 5.9:** *Parámetros encontrados para Secuencia 3.*

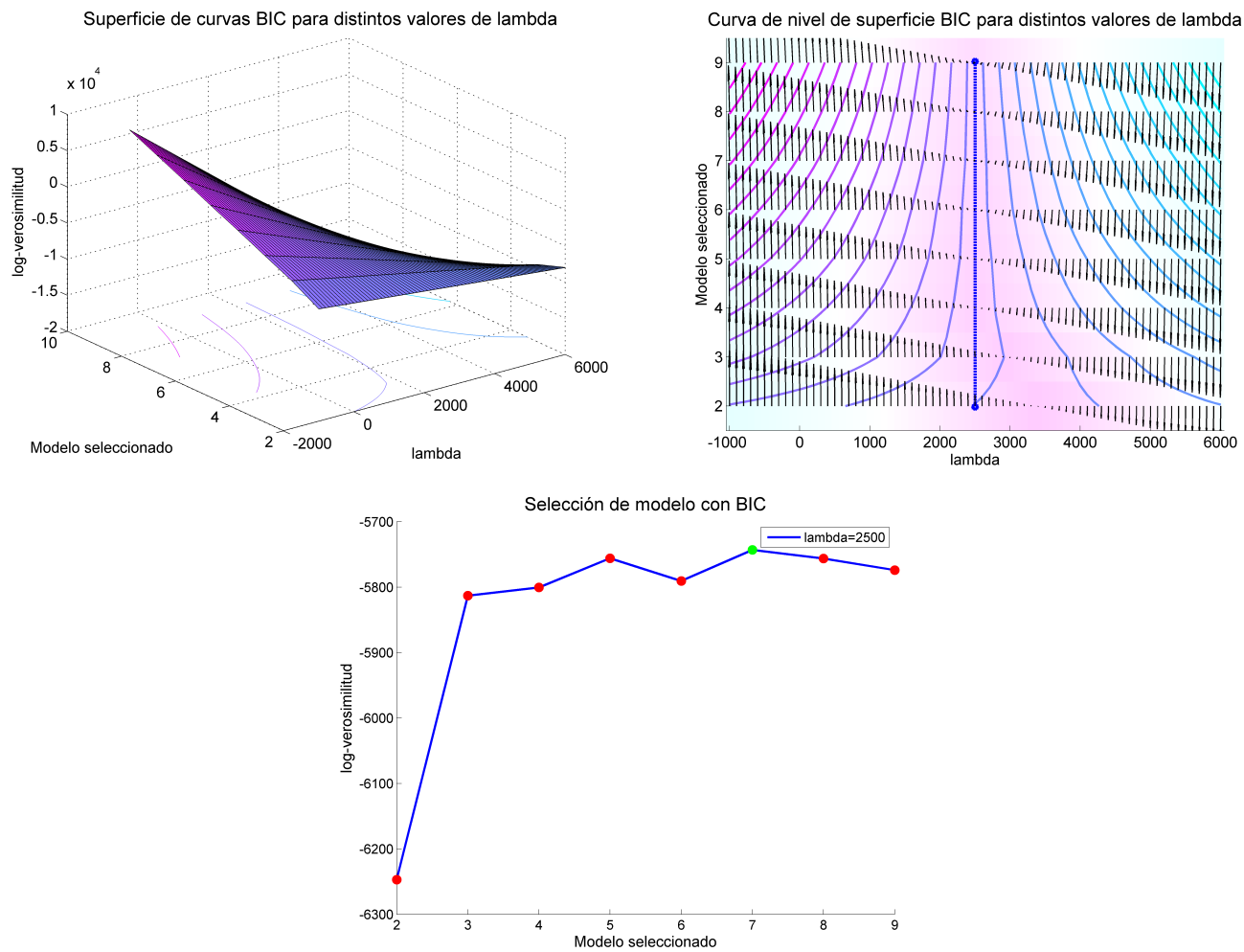


Figura 5.10: Superficie y curva de nivel BIC para Secuencia 3.

Descripción.

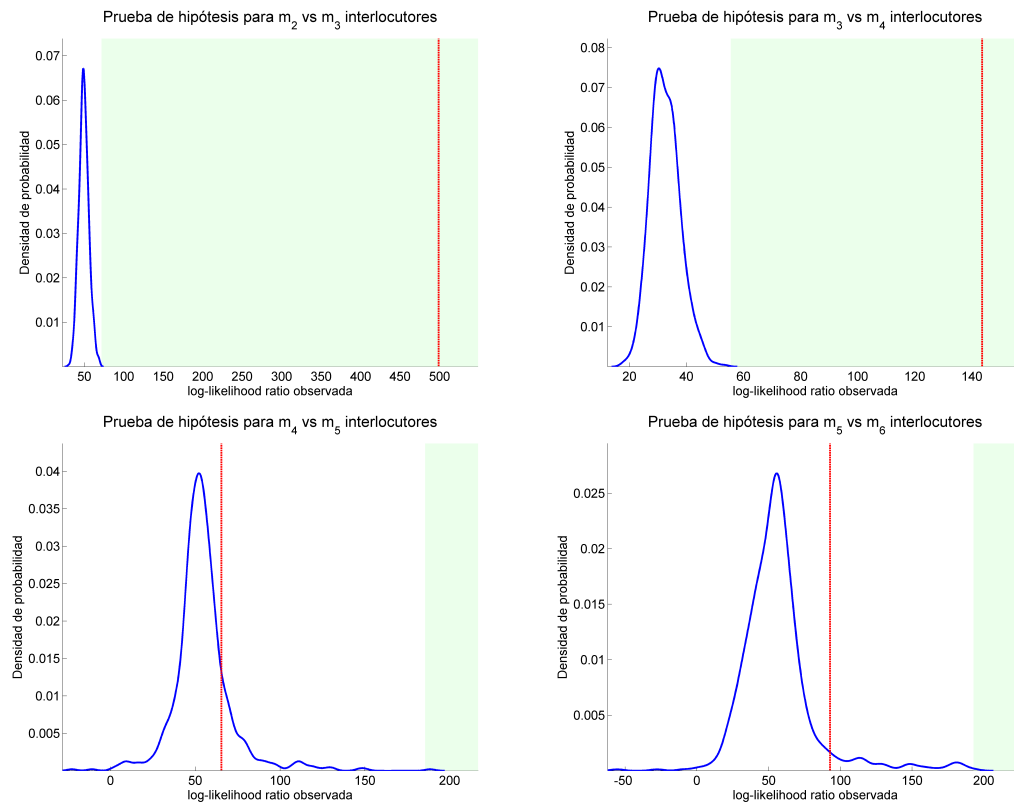


Figura 5.11: Pruebas de hipótesis con bootstrap para Secuencia 3.

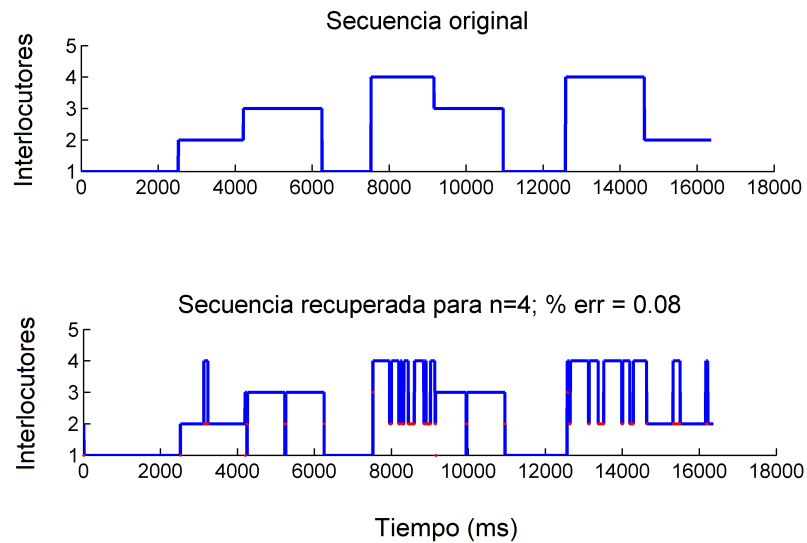


Figura 5.12: Secuencias encontradas para Prueba 3.

Secuencia 4: Manuel Acuña

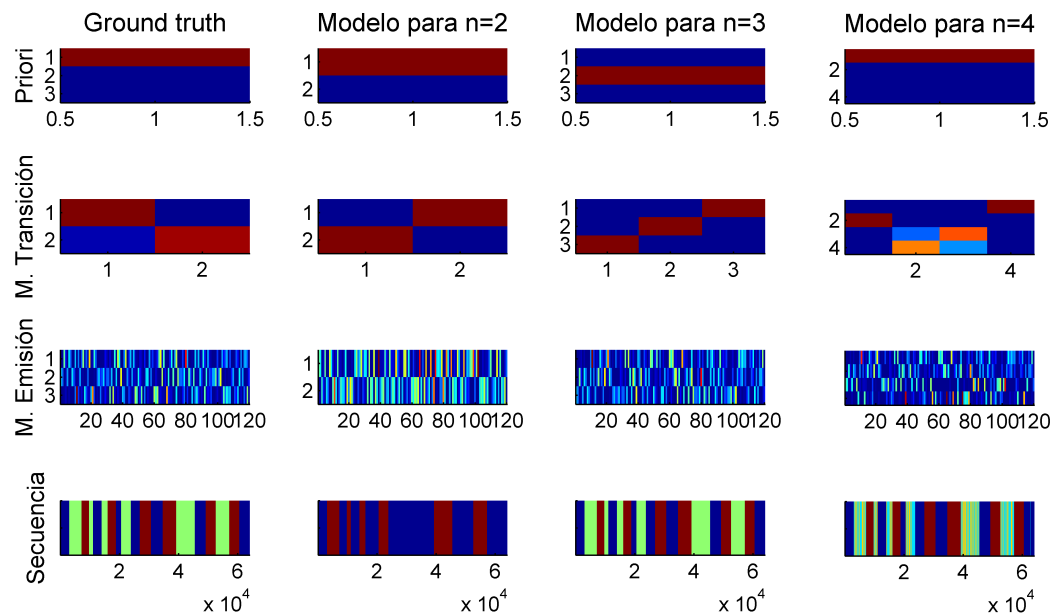


Figura 5.13: *Parámetros encontrados para Secuencia 4.*

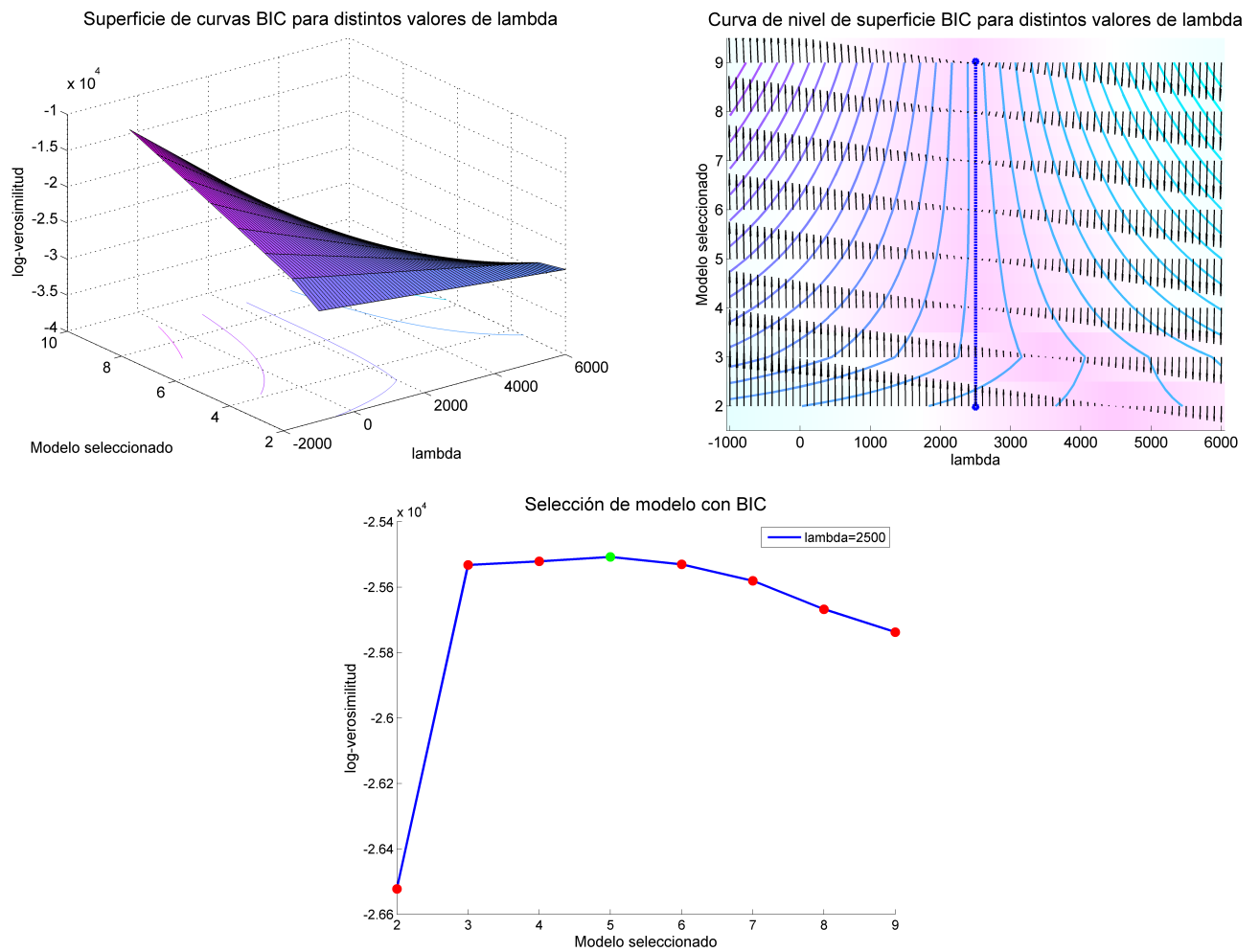


Figura 5.14: Superficie y curva de nivel BIC para Secuencia 4.
Descripción.

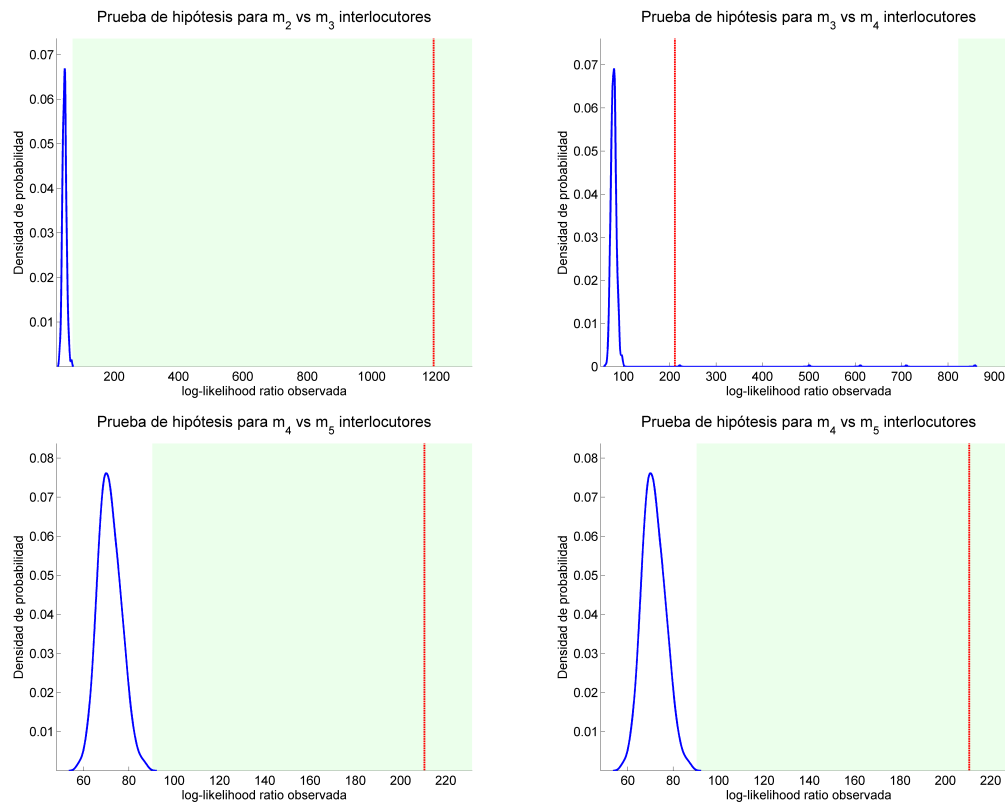


Figura 5.15: Pruebas de hipótesis con bootstrap para Secuencia 4.

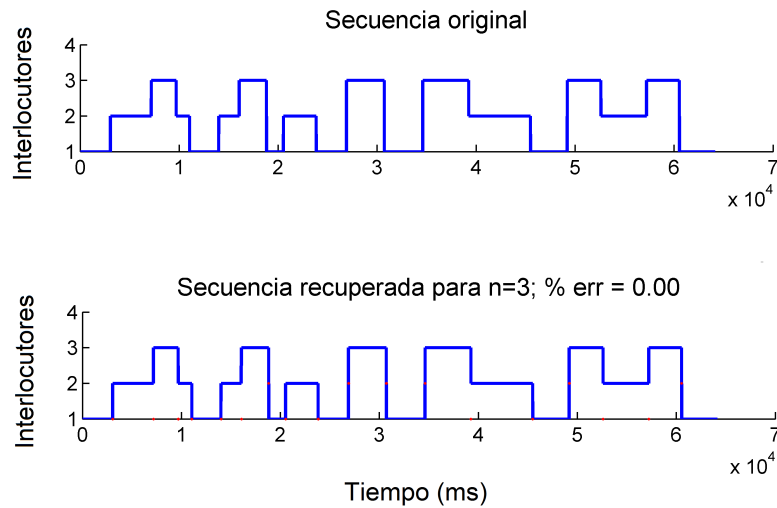
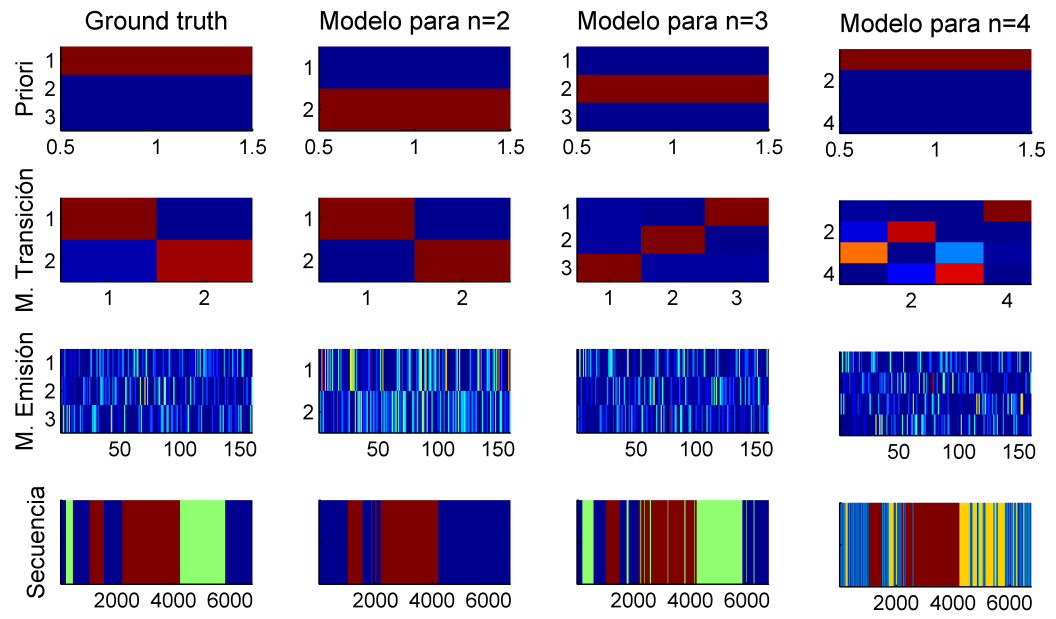
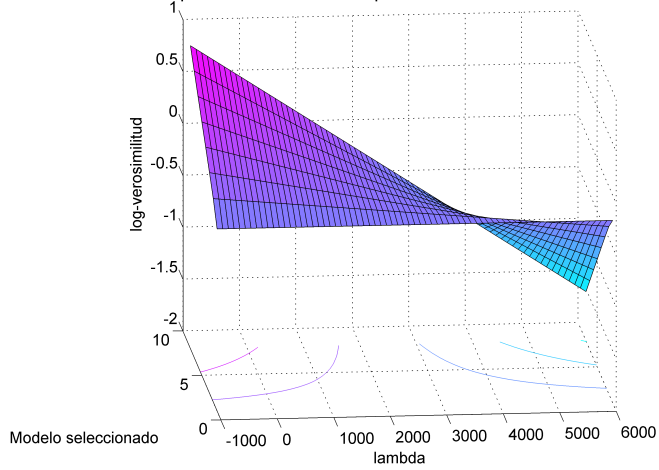


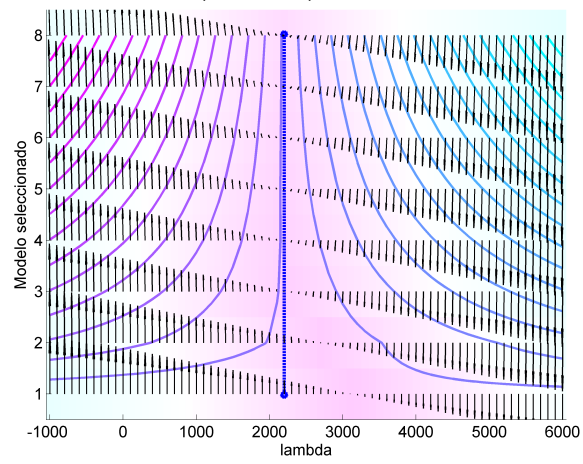
Figura 5.16: Secuencias encontradas para Prueba 4.

Secuencia 5: Calderón de la Barca**Figura 5.17:** *Parámetros encontrados para Secuencia 5.*

Superficie de curvas BIC para distintos valores de lambda



Curva de nivel de superficie BIC para distintos valores de lambda



Selección de modelo con BIC

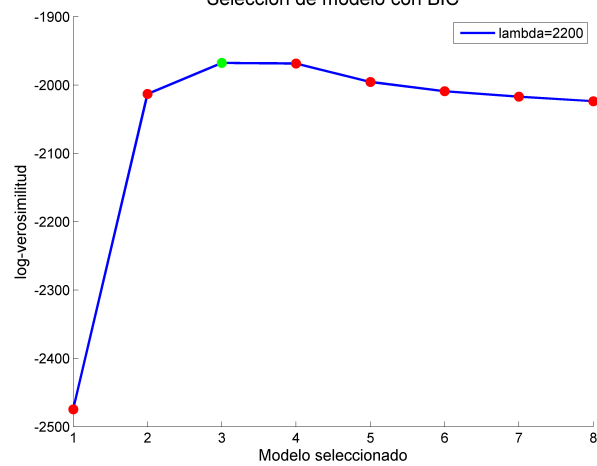


Figura 5.18: Superficie y curva de nivel BIC para Secuencia 5.
Descripción.

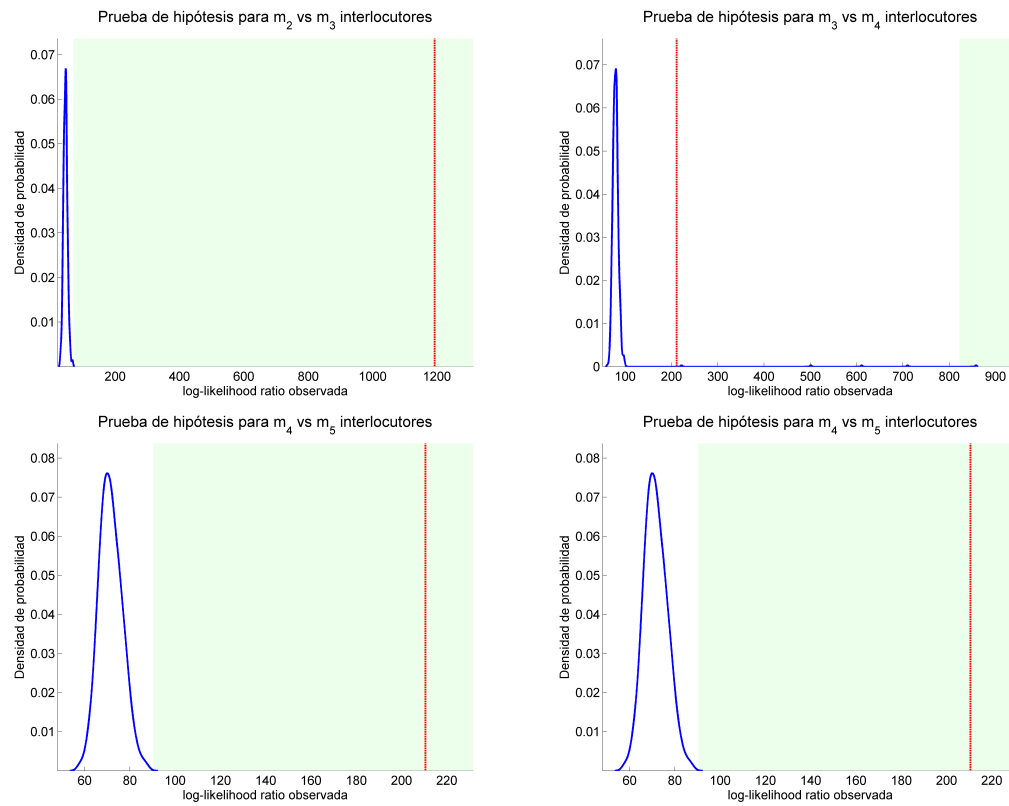


Figura 5.19: Pruebas de hipótesis con bootstrap para Secuencia 5.

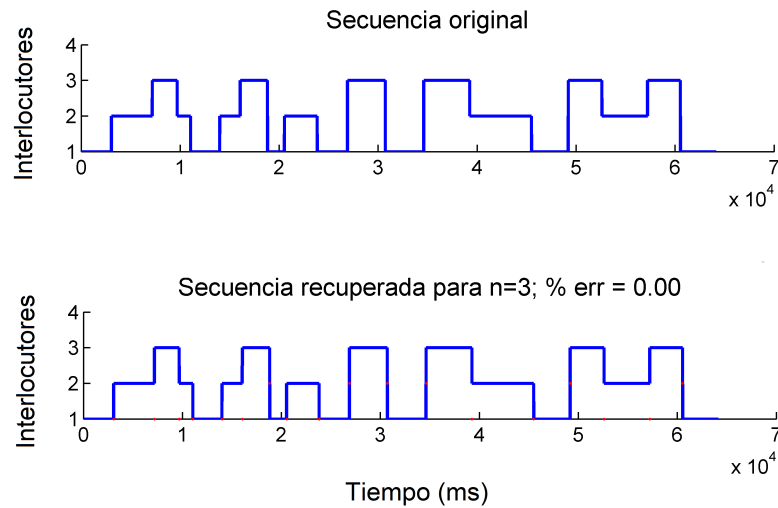


Figura 5.20: Secuencias encontradas para Prueba 5.

Secuencia 6: Andrew Lloyd Webber

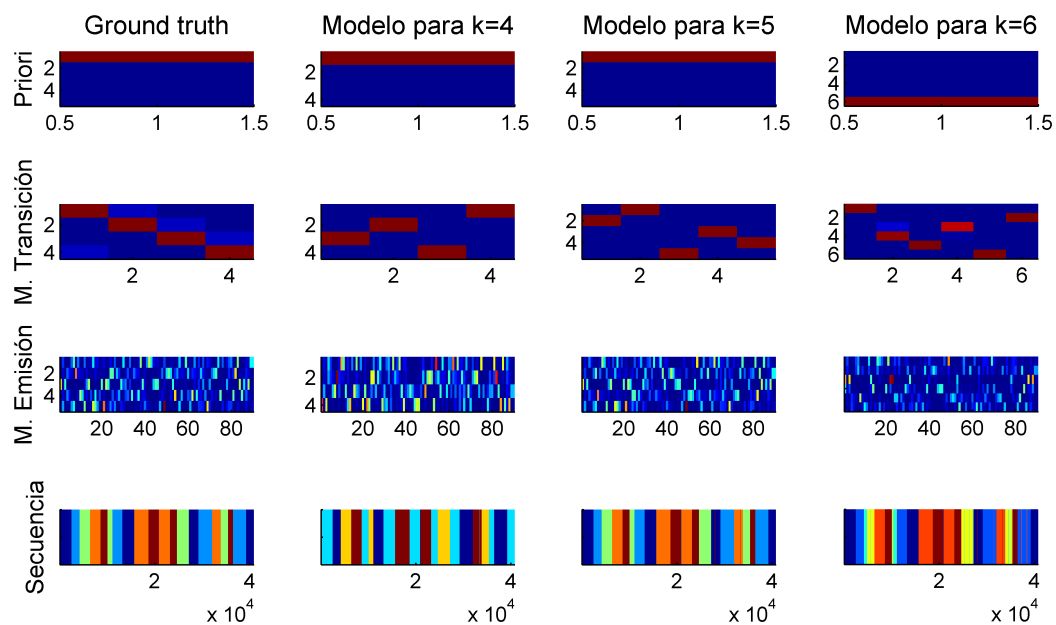


Figura 5.21: *Parámetros encontrados para Secuencia 6.*

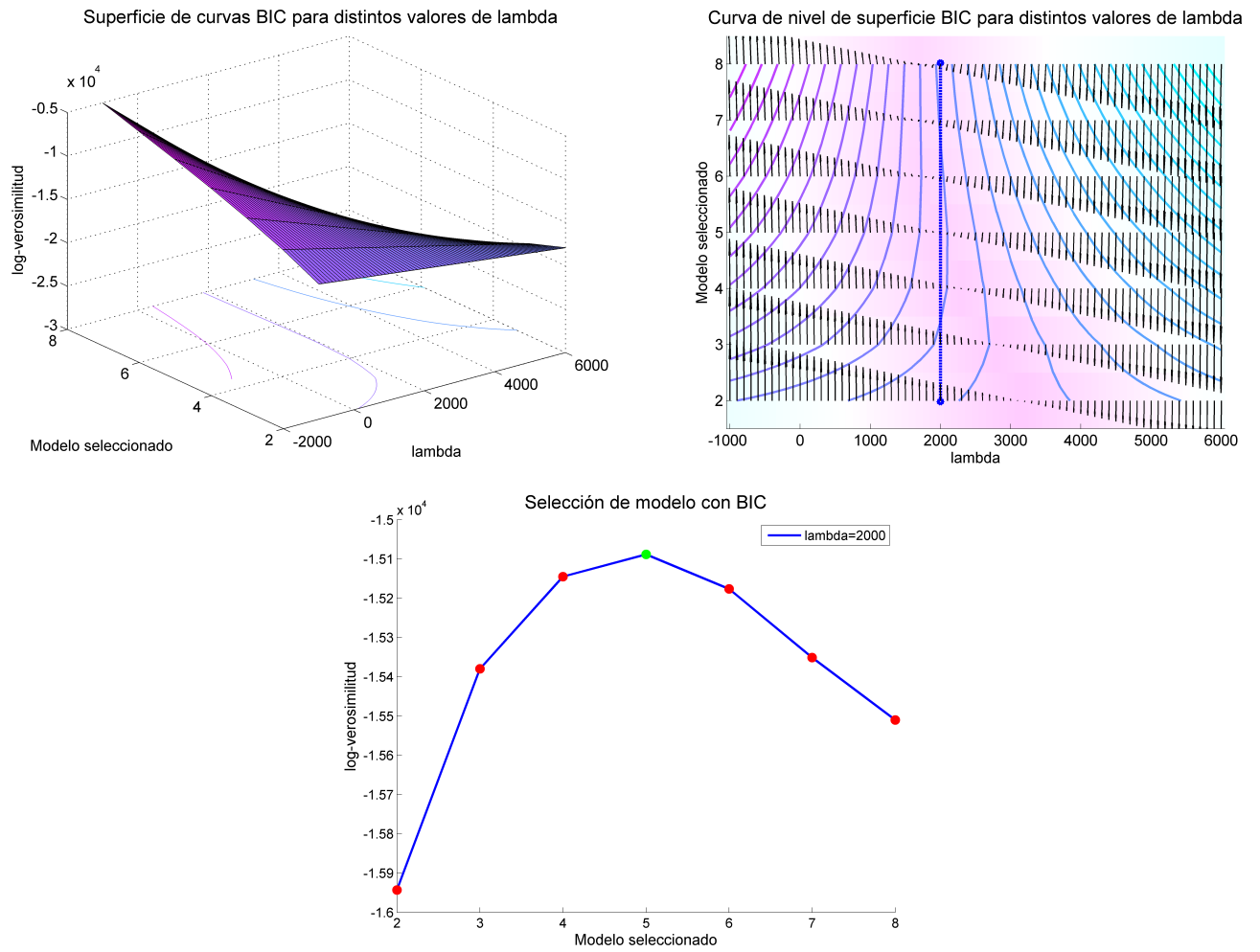


Figura 5.22: Superficie y curva de nivel BIC para Secuencia 6.
Descripción.

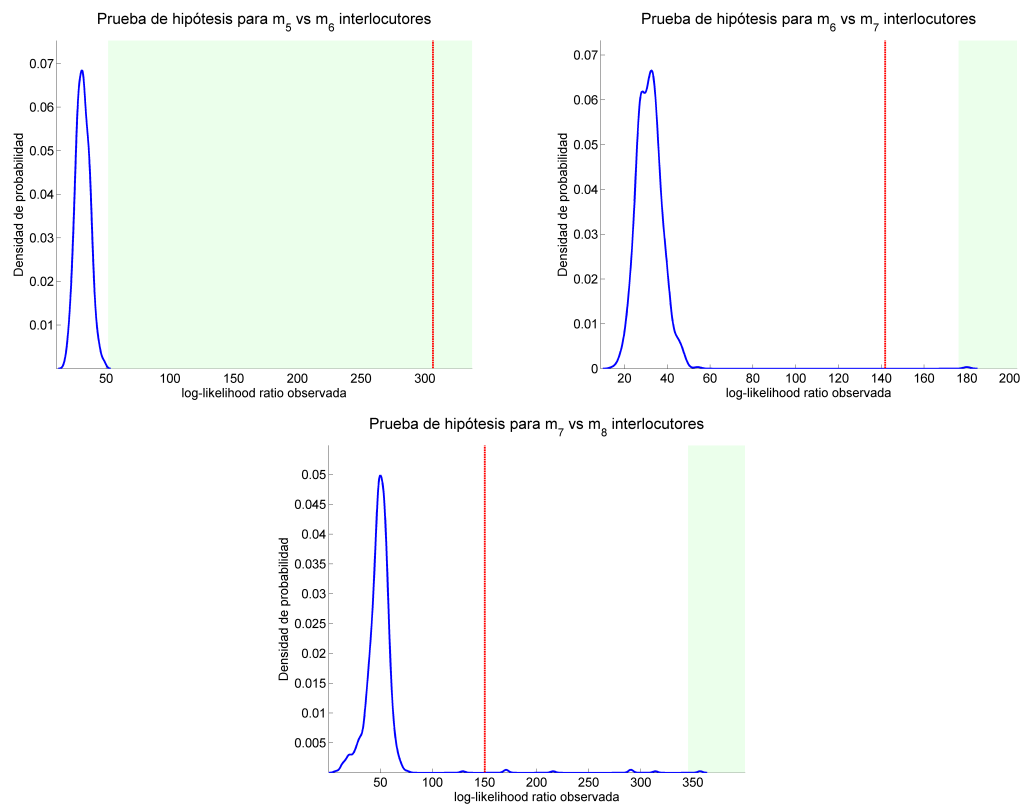


Figura 5.23: Pruebas de hipótesis con bootstrap para Secuencia 6.

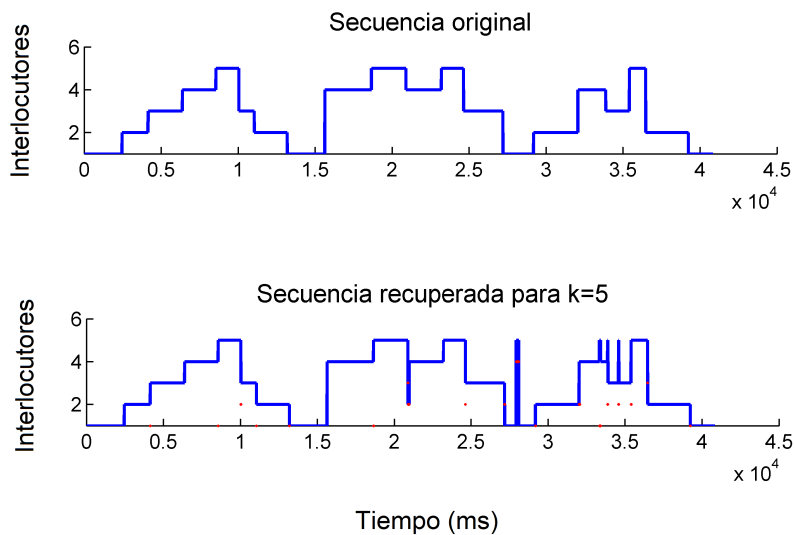


Figura 5.24: Secuencias encontradas para Prueba 6.

