



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**MODELAMIENTO Y CARACTERIZACIÓN DE CURVAS DE LUZ
CUASI-PERIÓDICAS UTILIZANDO MODELOS DE
NEUROPERCOLACIÓN**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA
INGENIERÍA, MENCIÓN ELÉCTRICA

CATALINA MARÍA PAZ ELZO VERA

PROFESOR GUÍA:
PABLO ESTÉVEZ VALENCIA

MIEMBROS DE LA COMISIÓN:
JORGE SILVA SÁNCHEZ
FRANCISCO FÖRSTER BURÓN
SUSANA EYHERAMENDY DUERR

SANTIAGO DE CHILE

2016

RESUMEN DE TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA ING.,
MENCIÓN ELÉCTRICA Y AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO.

POR: CATALINA MARÍA PAZ ELZO VERA
PROFESOR GUÍA: PABLO ESTEVEZ VALENCIA

FECHA: 14/11/2016

MODELAMIENTO Y CARACTERIZACIÓN DE CURVAS DE LUZ CUASI-PERIÓDICAS UTILIZANDO MODELOS DE NEUROPERCOLACIÓN

Los sistemas cuasi-periódicos han sido estudiados en el contexto de sistemas dinámicos, como también en diversas áreas en las cuáles estos fenómenos son observables. Tal es el caso de la astronomía. El caso de estudio presentado en este trabajo corresponde a las estrellas Gamma Doradus, las que debido a sus características físicas poseen entre 1 a 5 períodos principales de pulsación, otorgándoles características periódicas y cuasi-periódicas al ser observadas en el espectro visible. En el estudio de estas estrellas, la correcta determinación de sus periodos y de la distancia entre ellos es de vital importancia, debido a que se relacionan con sus parámetros físicos, los cuales pueden llegar a ser inferidos. Por tal motivo se espera que estos modelos ayuden en la tarea de caracterización de la pulsación, mediante la descomposición de una curva de luz en señales de menor complejidad.

La presente tesis se enfoca a la modelación y análisis de curvas de luz de estrellas Gamma Doradus mediante modelos de Neuropercolación. Neuropercolación es una familia de modelos estocásticos basados en la teoría de Autómata Celular Probabilístico en grillas y grafos aleatorios, inspirados en la dinámica de poblaciones neuronales. Estos modelos pueden ser utilizados para construir series de tiempo discretas, cuyos parámetros son las probabilidades que definen su dinámica. Se propone obtener tales parámetros para modelar una serie de tiempo mediante Optimización por Enjambre de Partículas (PSO), un método evolutivo de optimización basado en poblaciones.

La metodología propuesta se aplicó tanto en señales sintéticas como a curvas de luz reales de estrellas Gamma Doradus. Dentro de la señales sintéticas se estudió la capacidad de modelación de una señal de Amplitud Modulada, tanto en presencia de datos perdidos como de ruido blanco aditivo en distintos niveles. Para todos los casos se obtuvieron modelos representativos de la dinámica del proceso. Posteriormente, con modelos de estrellas Gamma Doradus se observa que, con ciertos modelos, es posible recuperar toda la base de frecuencias con la que se construyen las curvas de luz. Esto entrega ventajas comparativas con la búsqueda de frecuencias directamente sobre la señal, además de mejorar la estimación del espaciamiento entre periodos. Al analizar ese método en curvas de estrellas Gamma Doradus obtenidas por el proyecto Kepler de la NASA es posible observar que si bien los modelos no obtienen ajustes perfectos de Error Cuadrático Medio, algunas de las curvas si llegan a ser modeladas con una alta precisión y bajos residuos. Otros casos, en señales con mayores componentes espectrales, la modelación no obtiene tan buenos resultados, y los residuos, aún en los mejores casos, presentan estructuras que podrían llegar a ser modeladas.

Considerando la variabilidad intrínseca de resultados debido a la aleatoriedad de PSO y sus poblaciones iniciales, se concluye que es posible obtener modelos de señales cuasi-periódicas con gran exactitud, obteniendo las frecuencias principales del fenómeno. En los casos de modelos con menor ajuste la presencia de óptimos locales dificulta la convergencia hacia el óptimo global. Para los casos de estudio donde los errores fueron más altos, es posible que una mayor cantidad de datos ayude a la obtención de mejores modelos. Este trabajo presenta muchas opciones de extensiones, tales como la disminución de los tiempos de cómputo, mejoras del proceso de optimización, potencialidades en interpolación y predicción, y uso en series de tiempo con muestreos no-uniformes.

A mis abuelos.

Agradecimientos

Dado que la finalización de este trabajo de tesis representa el término de una etapa muy importante (y extensa) de mi vida, tengo mucha gente a quien agradecer por ser parte de este proceso.

Primero, por supuesto, tengo que agradecer a mi familia, quienes me han apoyado incondicionalmente durante toda mi vida. Así que gracias mamá, papá, Ignacio y Panchi por estar conmigo y tenerme más fe que la que yo me tengo. También, tengo que agradecer a mis abuelos, que aunque ya no estén acá siempre fueron una gran influencia en mi vida. Me hubiera gustado que pudieran estar presentes en estos momentos. Y además gracias a tíos, primos y familia extendida, especialmente a mi tía Cecilia y a la Cata.

También tengo mucho que agradecer a mis amigos. En primer lugar a mis amigas del colegio: Fafa, Panchu, Maca, Wii, Pily, Gene, Clau, Tacha, Jechu; con quienes he compartido gran parte de mi vida desde los primeros años de colegio. Luego, a los cabros de la U, especialmente al Pato, Toño, Sabaj, a quienes conocí muy pronto en la universidad; pero también a muchos otros de eléctrica, Leo, Tomi, Bachi, Jauregui; y a los vecinos del lab de control, Raulito, Victor y Caro. Y por supuesto, también a la Romi: yo creo que las cervezas en el Shamrock tuvieron una gran influencia en esta tesis.

Este trabajo definitivamente no habría sido posible sin el apoyo de mi profe guía, Pablo Estévez, y de la colaboración con Robert Kozma: muchas gracias a ambos por confiar en mí. Igualmente, gracias a toda la gente del Laboratorio de Inteligencia Computacional por la buena onda y los comentarios que ayudaron al desarrollo de esta tesis.

Seguro que me olvido de mucha gente que merece un lugar en esta página, pero tienen que saber que estoy eternamente agradecida por toda la gente que me ha ayudado a llegar hasta acá. Gracias a todos por el ánimo, la paciencia y el cariño.

Tabla de Contenido

1. Introducción	1
1.1. Hipótesis	5
1.2. Objetivos	5
1.2.1. Objetivo General	5
1.2.2. Objetivos Específicos	5
1.3. Estructura de la tesis	6
2. Antecedentes	7
2.1. Cuasi-periodicidad	7
2.2. Caso de estudio: Estrellas Gamma Doradus	9
2.2.1. Características físicas de la variabilidad en estrellas	10
2.2.2. Modelos para estrellas con variaciones gravitacionales	11
2.2.3. Factores que alteran la variabilidad estelar	12
2.2.4. Métodos de determinación de períodos de pulsación estelar	14
2.3. Métodos de análisis de frecuencia	16
2.3.1. Análisis espectral de Fourier	17
2.3.2. Periodograma de Lomb-Scargle	17
2.3.3. Otros métodos de análisis de frecuencia	19
2.4. Autómata celular	20
2.5. Percolación	21
2.6. Conjuntos K de Freeman	22
2.7. Neuropercolación	24
2.7.1. Ecuaciones del modelo <i>Mean Fields</i>	27
2.8. Optimización por enjambre de partículas	30
3. Metodología	31
3.1. Modelos	33
3.1.1. Modelos planos	33
3.1.2. Modelos esféricos	33
3.2. Métodos de estimación de parámetros	34
3.2.1. Optimización por enjambre de partículas	34
3.2.2. Manejo de restricciones	35
3.2.3. Creación de poblaciones iniciales	36
3.2.4. Funciones de fitness	37
3.3. Datos a utilizar	38
3.3.1. Datos sintéticos	38

3.3.2. Datos del proyecto Kepler	39
3.4. Herramientas de análisis de resultados	43
3.5. Programación	43
4. Resultados	44
4.1. Pruebas realizadas	44
4.2. Señal AM Sintética	45
4.2.1. Robustez ante datos perdidos	46
4.2.2. Robustez al ruido	49
4.3. Simulaciones de curvas de luz de estrellas Gamma Doradus	51
4.4. Curvas de luz reales de estrellas Gamma Doradus	55
4.4.1. Gamma Doradus KIC 9240041	55
4.4.2. Gamma Doradus KIC 10224094	57
4.4.3. Gamma Doradus KIC 11145123	58
5. Análisis y discusión de resultados	61
5.1. Señales Sintéticas	61
5.1.1. Señales de Amplitud Modulada	61
5.1.2. Curva de luz simulada de estrella Gamma Doradus	66
5.2. Curvas de luz reales de estrellas Gamma Doradus	69
5.2.1. Gamma Doradus KIC 9240041	69
5.2.2. Gamma Doradus KIC 10224094	71
5.2.3. Gamma Doradus KIC 11145123	72
5.3. Multiplicidad de los modelos	75
5.4. Capacidad de predicción	76
5.5. Transferencia de información e información <i>a priori</i>	77
5.6. Costo Computacional	78
6. Conclusiones	79
6.1. Trabajo Futuro	81
Bibliografía	83
7. Anexos	89
7.1. Ecuaciones de la variabilidad estelar	89
7.2. Códigos en MATLAB	90
7.2.1. Código para simulación de modelo KI	90
7.2.2. Código para simulación de modelo KII	91
7.2.3. Código para simulación de modelo KIII	92
7.3. Identificación de frecuencias en modelos de Neuropercolación	94

Índice de tablas

3.1. <i>Quarters</i> de datos del proyecto Kepler	39
4.1. MSE de los modelos de Neuropercolación para una señal AM sintética cuando se utiliza información incompleta para la estimación.	47
4.2. Períodos principales de los mejores modelos de Neuropercolación para una señal AM sintética con información incompleta.	47
4.3. MSE para modelos obtenidos con señales ruidosas.	49
4.4. Períodos principales de los modelos de Neuropercolación para una señal AM con ruido blanco.	50
4.5. Parámetros utilizados para la simulación de una curva de luz de estrella Gamma Doradus.	51
4.6. MSE de los modelos para Gamma Doradus simulada con amplitudes aleatorias.	53
4.7. MSE de los modelos para Gamma Doradus simulada con amplitudes uniformes.	54
4.8. MSE de los modelos para KIC 9240041	56
4.9. MSE de los modelos para KIC 10224094.	57
4.10. MSE de los modelos para KIC 11145123	60
5.1. Frecuencias identificadas por los mejores modelos para una curva de luz de Gamma Doradus simulada con amplitudes aleatorias.	69
5.2. Frecuencias identificadas por los mejores modelos para una curva de luz de Gamma Doradus simulada con amplitudes iguales.	69
5.3. MSE con datos de prueba en los modelos encontrados para KIC 9240041.	71
5.4. Frecuencias de los modos gravitacionales de la estrella KIC 11145123.	73

Índice de figuras

1.1. Ejemplo de un modelo de Neuropercolación.	3
2.1. Ejemplos de diferentes dinámicas observables en un espacio de fase	8
2.2. Coordenadas esféricas y nodos de variabilidad.	11
2.3. Espaciamiento de períodos para simulaciones de Gamma Doradus	14
2.4. Ejemplo de Bootstrap Percolation	21
2.5. Conjuntos K de Freeman.	25
2.6. Conjuntos K de Freeman utilizando Neuropercolación.	27
3.1. Metodología para el ajuste de un modelo de Neuropercolación.	32
3.2. Estructuras utilizadas para los modelos de Neuropercolación	33
3.3. Área de observación del telescopio espacial Kepler	40
3.4. Curvas de luz de estrellas Gamma Doradus después del pre-procesamiento.	42
4.1. Ejemplo de una señal de amplitud modulada simulada	45
4.2. Modelo y residuos al utilizar un 10% de los datos de una señal sintética AM.	46
4.3. Períodos de los mejores modelos de señal AM sintética con información incompleta.	48
4.4. Períodos de los peores modelos de señal AM sintética con información incompleta.	48
4.5. MSE de los modelos obtenidos al añadir ruido a una señal AM sintética.	50
4.6. Embedding para curva de luz simulada de Gamma Doradus.	52
4.7. Mejor modelo para Gamma Doradus simulada con amplitudes aleatorias, y sus residuos.	53
4.8. Mejor modelo para Gamma Doradus simulada con amplitudes uniformes, y sus residuos.	54
4.9. Embedding para la curva de luz de la estrella Gamma Doradus KIC 9240041.	55
4.10. Mejor modelo para KIC9240041	56
4.11. Embedding para la estrella Gamma Doradus KIC 10224094	57
4.12. Mejor modelo para KIC 10224094	58
4.13. Curva de luz de la estrella KIC 11145123, un híbrido entre Gamma Doradus y Delta Scuti.	59
4.14. Periodograma de Lomb Scargle de la curva de luz de la estrella KIC 11145123.	59
4.15. Mejor modelo para KIC 11145123 y residuos.	60
4.16. Periodograma de Lomb-Scargle de los residuos del mejor modelo para KIC 11145123.	60
5.1. Peor modelo encontrado para señal AM con el 10% de los datos.	61

5.2. Frecuencias encontradas en el mejor modelo de señal AM sin datos perdidos.	63
5.3. Histogramas de error en la modelación de señales con ruido.	64
5.4. Frecuencias encontradas en el mejor modelo de señal AM con SNR 30.	65
5.5. Modelos equivalentes para una señal AM	66
5.6. Frecuencias encontradas para curva de luz simulada con modelo de 4x4.	67
5.7. Espaciamiento de períodos para curva de luz simulada de amplitudes aleatorias con modelo de 4x4.	68
5.8. Frecuencias encontradas para curva de luz KIC 9240041.	70
5.9. Espaciamiento de períodos para KIC 9240041.	70
5.10. Frecuencias encontradas para curva de luz KIC 10224094.	72
5.11. Espaciamientos encontrados para curva de luz KIC 10224094.	72
5.12. Espectro completo de frecuencias encontradas para la curva de luz KIC 11145123.	74
5.13. Frecuencias asociadas a modos gravitacionales encontradas para la curva de luz KIC 11145123.	74
5.14. Ejemplo de multiplicidad de modelos.	75
5.15. Predicción para señal de Amplitud Modulada.	76
5.16. Predicción para curva de luz simulada de estrella Gamma Doradus.	76
7.1. Extracción de frecuencias en modelos de Neuropercolación.	94

1 Introducción

La cuasi-periodicidad es comúnmente definida de manera informal como un comportamiento *casi* periódico, sin embargo es un término técnico que representa la superposición de distintos períodos [1]. Una señal cuasi-periódica es aquella en la cual la evolución temporal puede ser descompuesta en múltiples componentes, cada una de ellas siendo periódica de manera independiente. Pese a ello, los períodos son suficientemente distintos como para no permitir la existencia de un período global T en la cual la dinámica se repite exactamente. De manera más formal, una solución para un sistema dinámico se dice cuasi-periódica con n frecuencias independientes ($n \geq 2$) si puede ser escrita como $F(x, t) = f(\omega_1 t, \omega_2 t, \dots, \omega_n t)$, donde $f(\cdot)$ es periódica de período 2π en cada uno de sus argumentos y el conjunto de frecuencias ($\omega_i > 0$) son racionalmente independientes ($\sum m_i \omega_i \neq 0$), es decir, inconmensurables [2] [3].

En astronomía hay diversos procesos que presentan comportamientos cuasi-periódicos, tales como los *Quasi-Periodic Oscillators* (QPO), observados en rayos X [4], mientras que otros tipos de comportamientos cuasi-periódicos y caóticos pueden observarse en estrellas. En particular, el proyecto de tesis aquí presentado utiliza como caso de estudio las estrellas Gamma Doradus [5, 6]. Éstas son estrellas pulsantes que presentan dinámicas periódicas y cuasiperiódicas, con presencia de uno o más períodos distintos. La pulsación de las estrellas puede ser observada mediante telescopios, como un cambio gradual en la luminosidad del objeto estelar. Las estrellas pulsantes son un área de estudio bastante amplia en el campo de la astronomía, dado que son útiles en una gran variedad de aspectos, desde determinar distancias estelares [7, 8] hasta obtener información sobre las características físicas de las estrellas [9]. Las estrellas Gama Doradus se encuentran en el último grupo, siendo importantes en el estudio de la Asterosismología, la ciencia que estudia la estructura interna de las estrellas pulsantes interpretando su espectro de frecuencias [10].

Es deseable poder estimar los múltiples períodos de las estrellas Gamma Doradus de forma precisa, debido a la información que tales datos pueden entregar sobre las propiedades estructurales de las estrellas. Sin embargo, existen problemas cuando se desea caracterizar un proceso cuasi-periódico en el espacio de frecuencias. Para series de tiempo cuasi-periódicas la Transformada de Fourier presenta máximos locales en frecuencias espurias, dado que se pueden observar máximos locales cada vez que la condición $\omega = \sum m_i \omega_i$ se cumple, para enteros m_i . Otros métodos de estimación de períodos, como Lomb-Scargle [11][12], que se basa en mínimos cuadrados, deben ser combinados con la verificación de expertos que conozcan las características esperables de este tipo de estrellas.

Adicionalmente la estimación de períodos asociados a la pulsación mediante fotometría es compleja debido a que los períodos de las órbitas de estas estrellas se encuentran en el mismo rango que sus períodos de pulsación. Esto provoca alteraciones en los períodos de oscilación que se observan en las series de tiempo de luminosidad de las estrellas, también llamadas curvas de luz. Actualmente gran parte del estudio de las estrellas Gamma Doradus no considera únicamente el análisis de curvas de luz, sino que utiliza herramientas adicionales, tales como la espectroscopía. Debido al costo asociado a la espectroscopía, es deseable minimizar la cantidad de observaciones de este tipo que son requeridas.

En esta tesis se propone la utilización de Neuropercolación para representar curvas de luz astronómicas cuasi-periódicas. Neuropercolación es una familia de modelos estocásticos basados en la teoría matemática de Autómata Celular Probabilístico [13] en grillas y grafos aleatorios. Neuropercolación, como su nombre lo sugiere, se encuentra motivado por las propiedades dinámicas y estructurales de las poblaciones neuronales [14, 15]. Esta familia de modelos presenta dinámicas inspiradas por los Conjuntos K propuestos por Walter Freeman en 1975 [16]. En estos modelos se define un grafo sobre una grilla con nodos determinados y conexiones entre ellos. Luego, un proceso dinámico es definido sobre el grafo como un proceso de percolación, es decir, un proceso donde un estado de activación se propaga a través de la grilla mediante reglas previamente definidas. El estado esperado de activación de los elementos de la grilla es la salida del modelo que se analiza en este trabajo. Dependiendo de los parámetros del modelo se pueden producir dinámicas que incluyen estabilidad, ciclos límites y oscilaciones cuasi-periódicas o caóticas. Estos modelos han sido utilizados exitosamente en modelamiento e interpretación de fenómenos observados experimentalmente en imágenes cerebrales, así como también en el desarrollo de aplicaciones ingenieriles para reconocimiento de patrones y control inteligente [17].

La familia de modelos de Neuropercolación, al estar basados en la dinámica de las poblaciones neuronales, presentan una jerarquía de configuraciones. Estas configuraciones van desde las más simples, que representan fenómenos básicos como la actividad de los receptores olfativos, hasta configuraciones complejas que pueden emular la actividad del bulbo olfatorio completo. Las estructuras de más bajo nivel son parte fundamental en la construcción de aquellas estructuras más complejas, siendo su interacción la que provee al sistema de la capacidad de desarrollar dinámicas caóticas o cuasi-periódicas. En particular, las estructuras utilizadas en este trabajo se encuentran compuestas por unidades que serán llamadas *osciladores*. Los modelos de Neuropercolación utilizados corresponden a arreglos ordenados de osciladores que interactúan entre ellos. Una explicación teórica detallada se encuentra en el Capítulo 2.

Un ejemplo se puede observar en la figura 1.1, que corresponde a una estructura de cuatro osciladores, y donde se detalla el comportamiento de la estructura entre las iteraciones 500 y 1500 del modelo. Cada oscilador se encuentra representado por un círculo, que presenta conexiones a un subconjunto de los demás osciladores. Las conexiones representan la influencia que un oscilador tiene sobre otro, la cual puede verse reflejada en un cambio de amplitud, de fase o de la composición de frecuencias del oscilador. Dentro de cada oscilador, hay neuronas o vértices de un grafo que cambian su estado de activación (activo-inactivo) constantemente. El valor esperado de estado de activación de todas las componentes de cada oscilador permite construir una serie de tiempo tal como las que se observan en la figura junto con los osciladores. El valor esperado de activación de la estructura completa, equivalente

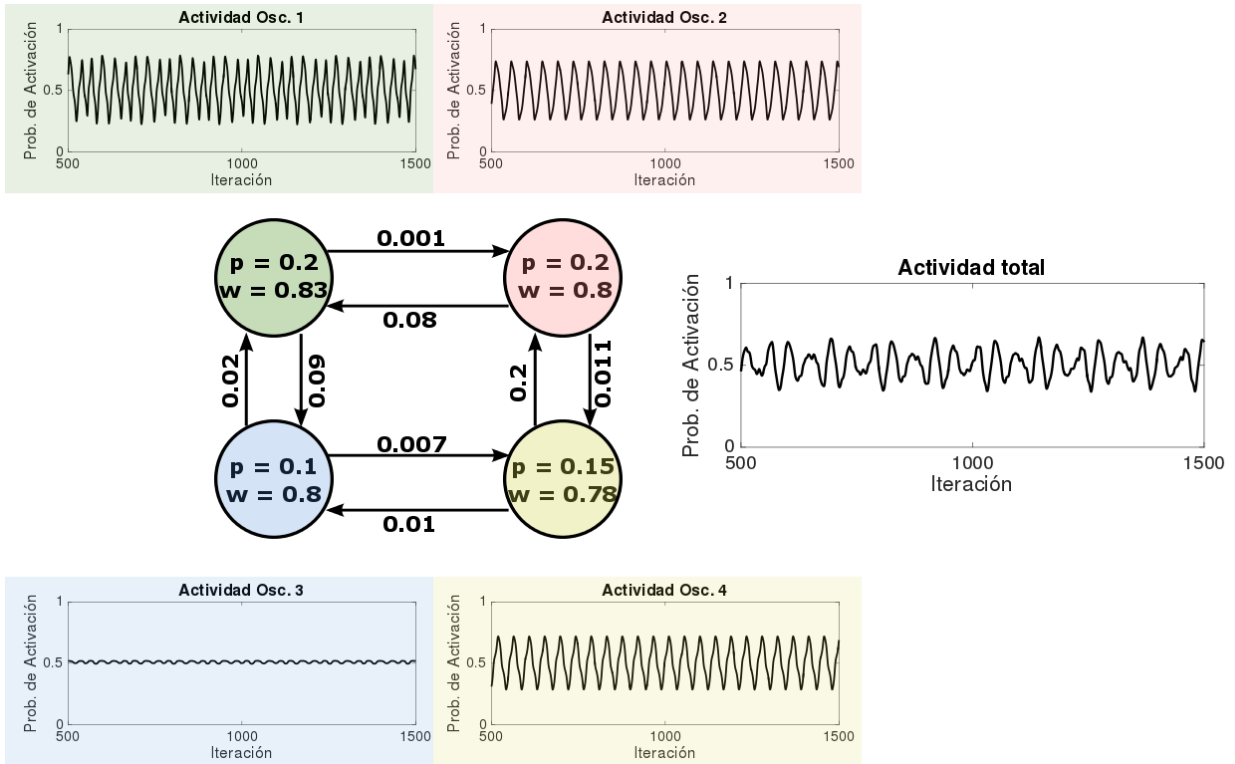


Figura 1.1: Ejemplo de un modelo de Neuropercolación con cuatro osciladores y sus correspondientes interconexiones. Cada gráfico muestra el estado de activación esperado para uno de los osciladores, más el estado de activación esperado total.

al promedio de activación de todos los osciladores, se transforma en otra serie de tiempo que puede representar dinámicas más complejas. Cada oscilador G_i tiene su comportamiento determinado por tres conjuntos de probabilidades: la probabilidad de seguir a la mayoría (ω_i), la probabilidad de inhibición (p_i) y las conexiones que pueda tener con otros osciladores G_j (c_{ji}). En la figura 1.1 se observa que ligeros cambios en los parámetros de los osciladores pueden llevar a dinámicas muy distintas. Por ejemplo, los osciladores 1 y 2 tienen frecuencias notoriamente diferentes, mientras que el oscilador 3 presenta amplitudes de oscilación muy pequeñas. La actividad total del modelo corresponde al promedio de las actividades de cada oscilador.

Las probabilidades utilizadas en los modelos de Neuropercolación son similares a las que se usan en un Autómata Celular Probabilístico, siendo éstos los parámetros que se deben ajustar para obtener un modelo adecuado para una señal en particular. Para encontrar los parámetros que representen un conjunto de datos específico en una estructura pre-definida, se propone utilizar Optimización por Enjambre de Partículas (PSO) [18, 19]. Este tipo de optimización basada en poblaciones resulta útil a la hora de optimizar grandes cantidades de parámetros en espacios donde no diferenciales, no convexos y con presencia de óptimos locales y globales. En particular el problema presentado puede presentar múltiples óptimos globales, los cuales se pueden obtener reordenando los osciladores, mientras sus conexiones relativas permanezcan intactas.

Debido a que el conjunto de parámetros es dependiente de la estructura del modelo, se deben analizar diversos modelos para encontrar aquel que entrega el mejor ajuste para cada curva de luz. Para medir el ajuste es posible utilizar medidas de comparación en tiempo (Error Cuadrático Medio, Correntropía) y frecuencia (comparación de espectros), aunque en este trabajo se utiliza únicamente el Error Cuadrático Medio como forma de comparación.

Esta tesis presenta dos motivaciones principales. Primero, el estudio de la capacidad que tienen los modelos de Neuropercolación para la modelación de señales cuasi-periódicas, en presencia de datos perdidos y ruido blanco aditivo. En segundo lugar, cómo este tipo de modelos pueden ser útiles para el análisis de series de tiempo complejas, mediante su descomposición en series de tiempo de menor complejidad. En el caso particular de las estrellas Gamma Doradus, es deseable la existencia de características que faciliten la estimación precisa de los distintos períodos de oscilación, que pueden ser difíciles de hallar con técnicas clásicas de análisis espectral.

Para abordar todos los puntos de interés, la metodología planteada es probada tanto para señales sintéticas como para curvas de luz reales de estrellas Gamma Doradus. Los datos sintéticos facilitan el estudio sobre los efectos que produce la falta de datos o el ruido, dado que son factores que se pueden controlar. Adicionalmente, si se conocen las componentes espectrales de una determinada señal, se pueden comparar éstas con las que son representadas por el modelo, de manera de estudiar la completitud del espectro obtenido al modelar la señal. El aplicar la metodología a señales reales permite la exploración de beneficios de los modelos de Neuropercolación en cuanto a la representación espectral, mediante la comparación de la componentes de frecuencia encontradas al analizar directamente la señal original versus analizar individualmente las componentes individuales del modelo (osciladores).

Entre las series de tiempo sintéticas a analizar se encuentra una señal de amplitud modulada con frecuencias inconmensurables, por lo que presenta características cuasi-periódicas, y simulaciones de curvas de luz de estrellas Gamma Doradus obtenidas de la literatura [20]. La primera señal es utilizada para evaluar cualitativamente el comportamiento de la metodología y la capacidad de modelamiento de Neuropercolación cuando la calidad de los datos se ve reducida, ya sea por pérdida de información o por ruido blanco aditivo.

Las señales reales se obtuvieron del proyecto Kepler [21], un telescopio espacial cuya misión principal es el descubrimiento de exoplanetas, o planetas fuera del Sistema Solar, pero que adicionalmente ha entregado información valiosa a diversas áreas de estudio en Astronomía y Astrofísica. En el caso particular de este trabajo, los datos de Kepler se ajustan a los requerimientos del estudio, dado su muestreo regular y bajo ruido. El proyecto Kepler posee información de 199 estrellas Gamma Doradus, según la clasificación publicada por el *Strasbourg Astronomical Data Center* (CDS) [22].

1.1. Hipótesis

Este trabajo busca probar empíricamente las siguientes hipótesis:

- Los modelos de Neuropercolación pueden representar y modelar series de tiempo dinámicas cuasi-periódicas.
- Los modelos de Neuropercolación permiten realizar un análisis en el dominio de la frecuencia de los osciladores que componen el modelo en lugar de la señal completa. Se desea comprobar que este tipo de análisis es ventajoso en señales con múltiples componentes espectrales.
- El método de optimización mediante poblaciones llamado Optimización por Enjambre de Partículas es útil para ajustar los parámetros de un modelo de Neuropercolación a una serie de tiempo dada.

1.2. Objetivos

1.2.1. Objetivo General

Modelar y analizar señales cuasi-periódicas sintéticas y curvas de luz reales mediante Neuropercolación.

1.2.2. Objetivos Específicos

- Modelar series de tiempo simuladas (Amplitud modulada y señales cuasi-periódicas) usando Neuropercolación. Analizar la calidad de los modelos en tiempo y frecuencia.
- Modelar curvas de luz con datos perdidos y distintos niveles de ruido, analizando el efecto que tiene la degradación de la información en el proceso de modelación.
- Modelar curvas de luz reales de objetos estelares cuasi-periódicos (Gamma Doradus).
- Caracterizar los modelos encontrados mediante las frecuencias principales asociadas a los osciladores de los modelos de Neuropercolación.
- Analizar los osciladores de los modelos obtenidos. Comparar las frecuencias obtenidas con las usadas para generar los datos en casos simulados, y aquellas reportadas en la literatura para casos de estrellas reales.

1.3. Estructura de la tesis

En el capítulo 2 se abordan los antecedentes requeridos para el desarrollo y comprensión de este trabajo de Tesis. Se define el concepto de Cuasi-Periodicidad, base para comprender la motivación y objetivos de este trabajo. Posteriormente se presentan detalles sobre el caso de estudio astronómico: las estrellas Gamma Doradus, explicando las razones de la cuasi-periodicidad observable en sus curvas de luz y la importancia que tiene la obtención precisa de sus componentes de frecuencia. Debido a que la obtención de espectros de frecuencia es un tema relevante para el análisis de resultados, se introducen los conceptos claves y los métodos utilizados. Luego, se explica en detalle el origen, propiedades y ecuaciones relacionadas a la familia de modelos de Neuropercolación. Finalmente se presenta el método de Optimización por Enjambre de Partículas, herramienta utilizada en la estimación de modelos.

En el capítulo 3 se presenta la metodología propuesta para resolver el problema planteado: la modelación de curvas de luz mediante modelos de Neuropercolación. Se discuten los detalles del proceso, el tipo de modelos a utilizar y las características y parámetros del método de optimización. Posteriormente se describen los datos a utilizar, sintéticos y reales, y su preprocesamiento. Finalmente se discute la manera en que se analizan los resultados y se entregan detalles sobre la implementación computacional realizada.

En el capítulo 4 se muestran los resultados de las pruebas realizadas, tanto con los datos sintéticos como reales. Se realizan pruebas de diversos modelos repetidas en múltiples ocasiones, para estudiar el efecto que la aleatoriedad del método de optimización tiene en los resultados.

En el capítulo 5 se discuten los resultados obtenidos contrastándolos con los resultados esperados. Se analiza la estabilidad de los resultados, además de la capacidad de representación de frecuencias que presentan los modelos de Neuropercolación. Se analizan los beneficios y debilidades de la metodología planteada, discutiendo posibles mejoras y detalles a considerar. Finalmente se presentan las conclusiones a este trabajo y sus posibles extensiones futuras.

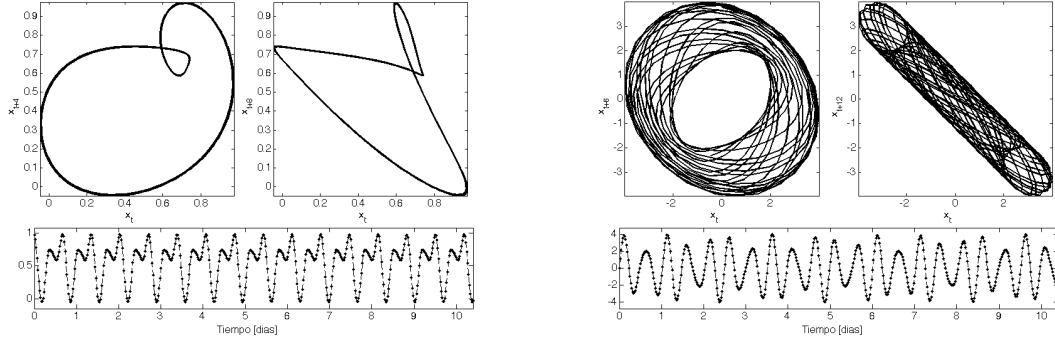
2 Antecedentes

2.1. Cuasi-periodicidad

El origen del estudio de sistemas dinámicos estuvo asociado a la física planetaria, aunque actualmente es un tema abarcado de forma interdisciplinaria. Los sistemas dinámicos son sistemas que evolucionan en el tiempo, y su evolución puede estar determinada por ecuaciones diferenciales (en tiempo continuo) o ecuaciones de diferencias (en tiempo discreto). Un sistema dinámico determinista puede ser transformado a una representación en variables de estado, las cuales definen una trayectoria de la evolución de un sistema. El espacio de estado se llama espacio de fase cuando es continuo, en el que se representan trayectorias continuas y suaves. Estas trayectorias pueden entregar una gran cantidad de información cualitativa asociada a la dinámica del sistema, la cual suele ser bastante difícil de obtener mediante el análisis directo de las ecuaciones diferenciales debido a su complejidad.

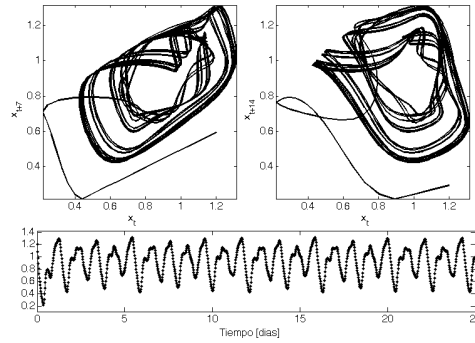
Algunas características que pueden ser obtenidas usando representaciones de espacio de fase tienen relación con el comportamiento de las soluciones del sistema, ya sea periódico, cuasi-periódico, caótico, atractor, etc. Las dinámicas periódicas son aquellas que repiten su comportamiento en intervalos fijos de tiempo T ; las cuasi-periódicas son aquellas que no presentan un comportamiento periódico, pero sí una estructura regular, la que puede observarse en un diagrama de fase con la forma de un hipertoroide. Los sistemas caóticos tienen como principal característica que pequeñas variaciones en los parámetros del sistema, o en la condición inicial pueden llevar a resultados divergentes exponencialmente. Todas estas dinámicas se ilustran en el dominio temporal y en su correspondiente diagrama de fase en la figura 2.1.

Según el Teorema de *Embedding* de Takens [23], los diagramas de fase pueden ser reconstruidos a partir de una única serie de tiempo. Esto se logra mediante la construcción de un vector de estados correspondientes a la serie de tiempo retardada en el tiempo, $[y(t), y(t + \tau), y(t + 2\tau), \dots, y(t + (M - 1)\tau)]$. Para construir este vector es necesario encontrar τ , que es el retardo en el tiempo, y M que es el tamaño del vector y determina la cantidad de dimensiones del espacio de fase. Los parámetros óptimos del *embedding* pueden ser obtenidos mediante diferentes técnicas. Una discusión extensa sobre tales métodos puede encontrarse en el libro *Nonlinear Time Series Analysis* de Kantz y Schreiber [1].



(a) Dinámica periódica

(b) Dinámica cuasi-periódica



(c) Dinámica caótica

Figura 2.1: Ejemplos de diferentes dinámicas observables en un espacio de fase. Cada dinámica se representa en espacios bidimensionales ($M = 2$), con retardos distintos. La figura 2.1(a) corresponde a un sistema con comportamiento periódico, donde las trayectorias son cerradas, formando ciclos límites. La figura 2.1(b) representa un sistema cuasi-periódico, donde los ciclos no se superponen de manera exacta, sino que se observan como un toroide. La figura 2.1(c) ilustra un sistema caótico, donde pequeños cambios en las condiciones iniciales conllevan trayectorias diametralmente diferentes.

Las oscilaciones cuasi-periódicas pueden aparecer cuando un oscilador es forzado por una entrada que es función del tiempo (oscilaciones cuasi-periódicas forzadas), o bien ser generadas intrínsecamente por un sistema no lineal no forzado (auto oscilaciones cuasi-periódicas) [3]. Los sistemas dinámicos pueden ser no-autónomos o autónomos, dependiendo de si la ecuación diferencial incluye explícitamente o no la variable independiente (usualmente el tiempo). El estudio de oscilaciones cuasi-periódicas, en el contexto de sistemas dinámicos, usualmente se reduce al sistema no-autónomo de ecuaciones diferenciales ordinarias de la ecuación 2.1, donde el lado derecho es cuasi-periódico con respecto a t ,

$$\frac{d\mathbf{x}}{dt} = F(t, \mathbf{x}). \quad (2.1)$$

Una señal cuasi-periódica es aquella en la cual su evolución temporal puede descomponerse en componentes periódicas, pero cuyos períodos no son congruentes. Por lo tanto no existe un período global T en el cuál la señal regresa exactamente al mismo punto de su trayectoria.

Matemáticamente, una solución se dice cuasi-periódica con n frecuencias independientes ($n \geq 2$) si puede ser escrita como

$$F(t, x) = f(\omega_1 t, \omega_2 t, \dots, \omega_n t), \quad (2.2)$$

donde $f(\phi_1, \dots, \phi_n)$ es periódica de período 2π en cada uno de sus argumentos y el conjunto de frecuencias ($\omega_i > 0$) son racionalmente independientes ($\sum m_i \omega_i \neq 0$). Este tipo de frecuencias se llaman *incommensurables*, es decir, no están relacionadas mediante números racionales [1]. Estas soluciones definen un toroide n -dimensional en el espacio de fase [2] [3].

De esta manera, el sistema cuasi-periódico de la ecuación 2.1 puede expresarse en variables de fase $\phi = (\phi_1, \dots, \phi_n)$, como

$$\frac{d\phi}{dt} = \omega \quad \frac{dx}{dt} = F(\phi, x), \quad (2.3)$$

donde $\omega = (\omega_1, \dots, \omega_n)$ es la base de frecuencias de la función cuasi-periódica.

En presencia de series de tiempo cuasi-periódicas la Transformada de Fourier usualmente presenta máximos locales en frecuencias distintas a las de los fenómenos oscilatorios reales, cuando se cumple la condición $\omega = \sum m_i \omega_i$ para enteros m_i arbitrarios. Debido a esto, puede resultar difícil discriminar comportamientos cuasi-periódicos de comportamientos caóticos únicamente por su espectro de frecuencia.

2.2. Caso de estudio: Estrellas Gamma Doradus

Las estrellas Gamma Doradus deben su nombre a la tercera estrella más brillante de la constelación de Dorado. La primera caracterización de su variabilidad fue realizada en 1963 por Cousins y Warren [24]. Posteriormente se definió que poseía dos períodos principales en 0.733 días y en 0.757 días. En esa época, al no existir una clasificación apropiada para esta estrella, fue etiquetada como una variable sin causa. Estudios posteriores entregaron mayor información sobre esta estrella y permitieron la clasificación de ésta y otras similares dentro de un mismo grupo.

Las estrellas Gamma Doradus son variables pulsantes, tipo de variabilidad que fue identificado por Balona et al. en 1994 [5]. En el año 1999 fueron definidas como un nuevo tipo de estrellas pulsantes por Kaye et al [6]. Recientemente, las estrellas Gamma Doradus han sido estudiadas en más profundidad, concluyendo que sus variaciones se deben principalmente a la interacción entre oscilaciones no-radiales, que presentan una influencia estabilizadora, y el envoltorio convectivo de la estrella, que tiene una influencia desestabilizadora [25].

Sus principales características, relacionadas a sus curvas de luz, son que presentan de 1 a 5 períodos principales y observables en telescopios terrestres, en un rango entre 0.3 y 3 días. En cuanto a sus características físicas, son estrellas jóvenes, cercanas a la secuencia principal.

Presentan entre 1.5 y 1.8 masas solares y períodos de rotación que pueden llegar a ser cercanos a sus períodos de oscilación, por lo que en ese sentido son osciladores relativamente rápidos. Esto produce alteraciones en los períodos observables [26].

Pese a que estas estrellas han sido caracterizadas de manera muy reciente, ya han sido utilizadas para el estudio de la Asterosismología [27]. Esta ciencia estudia la estructura interna de las estrellas pulsantes interpretando su espectro de frecuencia. Se han construido modelos que relacionan parámetros físicos de estrellas con características similares a las Gamma Doradus (gravedad, densidad, presión y radio), con los espaciamientos entre los períodos observables en sus pulsaciones [10]. Los telescopios espaciales como Kepler y CoRoT han sido esenciales en este aspecto, dado que entregan datos con una gran precisión, la que es necesaria para estos estudios.

2.2.1. Características físicas de la variabilidad en estrellas

Las estrellas, al ser cuerpos tridimensionales, tienen modos naturales de oscilación que presentan nodos en tres direcciones ortogonales, descritas como r (distancia al centro), θ (co-latitud) y φ (longitud), tal como se observa en la figura 2.2. Los nodos pueden visualizarse como cáscaras para r constante, conos para θ constante y planos para φ constante. En la mayoría de las estrellas pulsantes, grupo del cual las Gamma Doradus son parte, el eje de pulsación coincide con el eje de rotación. Las ecuaciones que gobiernan estas oscilaciones, en las direcciones previamente mencionadas, se encuentran detalladas en el anexo 7.1.

Las ecuaciones de variabilidad son determinadas por tres números cuánticos, n , l y m . Estos números están asociados al modo de oscilación. El parámetro n se relaciona con la cantidad de nodos radiales y es llamado el *overtone* del modo. El parámetro l es el grado del tono, y especifica el número de nodos en la superficie. Finalmente, m es el orden azimutal del modo, y su módulo $|m|$ especifica cuántos de los nodos superficiales son líneas de longitud. El signo de m señala si las ondas se trasladan en la dirección de la rotación o en contra de ésta. Los valores de m se encuentran en el rango entre $-l$ y l , por lo que existen $2l + 1$ modos por cada grado l [9].

Existen dos tipos de soluciones a la ecuación de movimiento para una estrella pulsante, lo cual lleva a dos modos de pulsación: *modos p* y *modos g*. En los modos p o de presión, la presión es la principal fuerza restauradora en una estrella perturbada del equilibrio. Estos modos corresponden a ondas acústicas con movimientos principalmente verticales. En el caso de los modos g, o gravitacionales, el empuje es la principal fuerza restauradora y los movimientos son principalmente horizontales [9]. En particular las estrellas Gamma Doradus pulsán en modos gravitacionales no-radiales.

El modo radial más simple corresponde al *dipolo*, con $l = 1$ y $m = 0$. En este modo, el ecuador es un nodo, y el hemisferio norte se expande mientras que el hemisferio sur se contrae, y viceversa. El centro de masa de la estrella no se desplaza con estos movimientos, por lo que las oscilaciones tipo dipolo son posibles. Las estrellas Gamma Doradus generalmente presentan modos con un alto valor de n , donde $n \gg l$. En adelante, éstos serán llamados modos gravitacionales de alto orden.

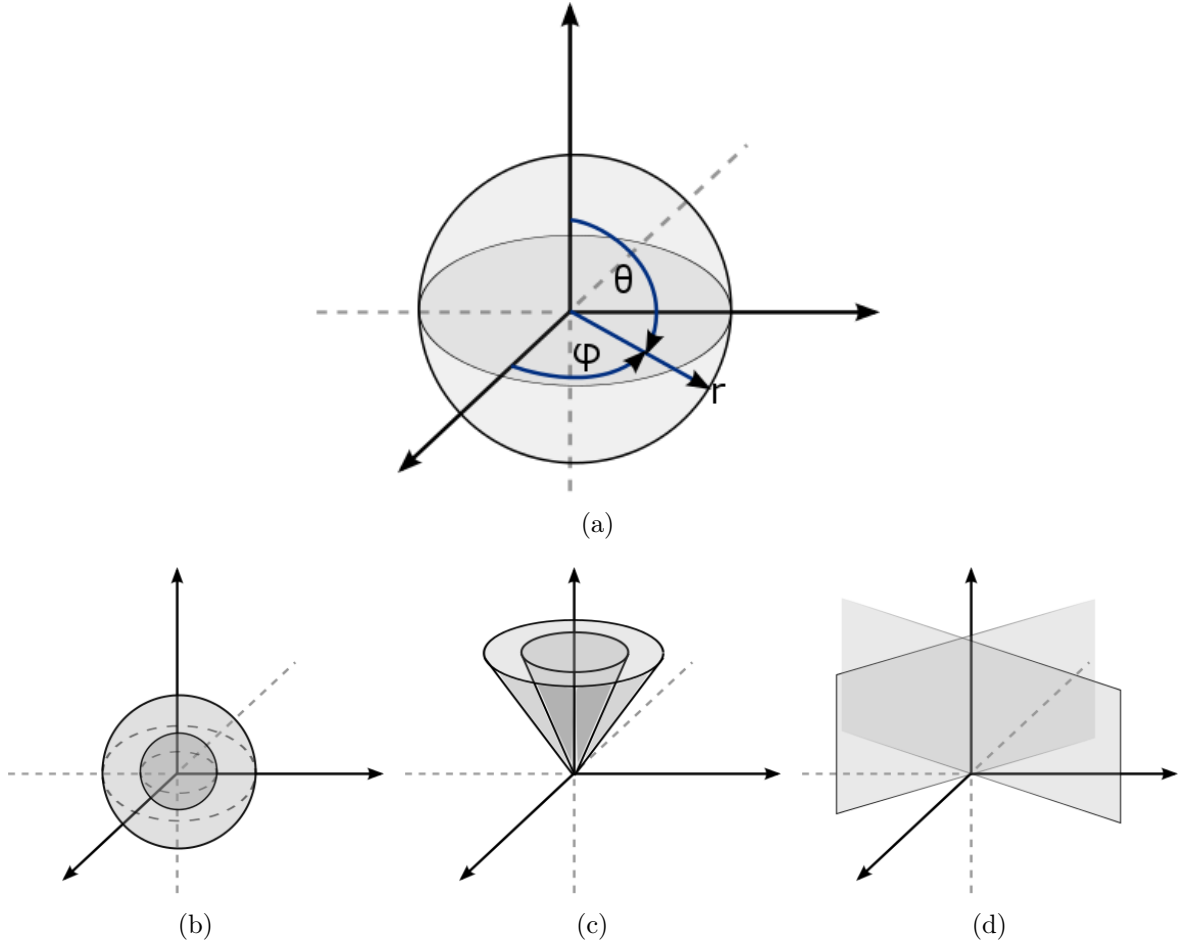


Figura 2.2: Coordenadas esféricas y los nodos de variabilidad que se obtienen para coordenadas constantes. El sistema de coordenadas se muestra en la figura 2.2(a). Cuando los radios (r) son constantes se forman superficies esféricas como las cáscaras de la figura 2.2(b). Colatitud constante (θ) se refleja en superficies con forma de cono, como las de la figura 2.2(c), mientras que se obtienen planos tales como los de la figura 2.2(d) al fijar un valor para la longitud (φ).

2.2.2. Modelos para estrellas con variaciones gravitacionales

Las propiedades de los espectros de modos gravitacionales de alto orden se encuentran determinadas por la estratificación de la materia en la estrella, la cual se encuentra descrita por la frecuencia de Brunt-Väisälä (N). Esta señala la frecuencia angular a la cual una parcela de fluido oscila en un ambiente estáticamente estable. La relación está dada por [10]

$$N^2 = g \left(\frac{1}{\Gamma_1} \frac{d \ln P}{dr} - \frac{d \ln \rho}{dr} \right) \propto \left(\nabla_{ad} - \nabla + \frac{\varphi}{\delta} \nabla_{\mu} \right), \quad (2.4)$$

donde g es la gravedad local, ρ es la densidad local, P es la presión local, r es el radio local, Γ_1 es el primer exponente adiabático, ∇_{ad} y ∇ son los gradientes adiabático y de temperatura estelar, ∇_{μ} es el gradiente promedio de peso molecular, y φ y δ son las derivadas parciales de

la densidad con respecto al peso molecular promedio (μ) y a la temperatura, respectivamente.

Mediante una aproximación asintótica de primer orden [28], es posible hacer una conexión entre el espectro de un modo gravitacional y las características físicas de la estrella. Según esta teoría, en una estrella con un núcleo convectivo y envoltorio radiativo, el periodo $P_{n,l}$ de un modo gravitacional de alto orden, $n \gg l$, y grado l se encuentra dado por:

$$P_{n,l} = \frac{\Pi_0}{\sqrt{l(l+1)}} (n + \alpha_{l,g}) \quad (2.5)$$

donde

$$\Pi_0 = 2\pi^2 \left(\int_{r_1}^{r_2} N \frac{dr}{r} \right)^{-1} \quad (2.6)$$

y $\alpha_{l,g}$ es un coeficiente que depende de los límites de la región, r es la distancia al centro estelar y N es la frecuencia de Brunt-Väisälä.

Por otro lado, en modos de alto orden, ignorando efectos rotacionales, existe una relación asintótica para modos de presión que señala que se encuentran aproximadamente equiespaciados en frecuencia. De la misma manera existe otra relación asintótica para modos gravitacionales que señala que éstos se encuentran equiespaciados en periodo [9]. Aquellos modos con el mismo grado l , y órdenes consecutivos n , se encuentran separados equidistantemente con respecto a su periodo, con el valor dado por

$$\Delta P_l = \frac{\Pi_0}{\sqrt{l(l+1)}}. \quad (2.7)$$

Sustituyendo 2.6 en 2.7 se obtiene el espaciado entre períodos:

$$\Delta P_l = P_{n+1,l} - P_{n,l} = \frac{2\pi^2}{\sqrt{l(l+1)} \int_{r_1}^{r_2} \frac{N}{r} dr}. \quad (2.8)$$

Es claro que fuertes supuestos son aplicados en la creación de estos modelos, dado que no tienen en cuenta la rotación de las estrellas. Éste es un factor altamente relevante, debido a que los períodos de rotación de las estrellas Gamma Doradus son similares a sus períodos de pulsación lo que provoca distorsiones en la luminosidad observada.

2.2.3. Factores que alteran la variabilidad estelar

El efecto que tiene la rotación en los períodos observables es un fenómeno que ha sido estudiado recientemente. Bouabid et al. (2013) [26] estudiaron el efecto de la fuerza de Coriolis para estrellas Gamma Doradus pulsando en diversos modos, utilizando aproximaciones acordes al problema. La aproximación empleada permite la separabilidad del problema por

medio de despreciar la componente de latitud en el vector de rotación de la fuerza de Coriolis. Esta aproximación es adecuada bajo los siguientes supuestos:

- El desplazamiento horizontal debe ser mucho más fuerte que el radial, y por lo tanto Coriolis no tiene un papel significativo en la componente radial de la ecuación de momentum.
- La estrella debe rotar lenta o moderadamente con respecto a su velocidad de rotación crítica, definida como

$$\Omega_c \approx \sqrt{\frac{8GM_*}{27R_*^3}} \quad (2.9)$$

donde M_* y R_* son la masa y radio de la estrella, respectivamente, y G es la constante de gravitación universal. Este supuesto permite obviar la fuerza centrífuga.

Al utilizar esta aproximación, no se deben realizar demasiados cambios en las ecuaciones ya descritas de los períodos. En la ecuación 2.5 se debe reemplazar el valor de l por \tilde{l} que corresponde al grado armónico efectivo introducido por Towsend [29]. Por lo tanto, la siguiente ecuación,

$$P_{co} = \frac{\Pi_0}{\sqrt{\tilde{l}(\tilde{l} + 1)}} (n + \alpha_{l,g}), \quad (2.10)$$

entrega el valor del periodo en un modo dado para un eje que rota en conjunto con la estrella. Para transformar este periodo a uno visto desde un eje inercial se debe aplicar la siguiente relación, donde P_Ω es el periodo rotacional de la estrella:

$$P_{in} = \frac{P_{co}}{1 - m \frac{P_{co}}{P_\Omega}}. \quad (2.11)$$

La fuerza de Coriolis afecta de manera distinta si la estrella pulsa en modo progrado, donde $m > 0$ y las ondas se propagan en el sentido de la rotación, o retrógrado, en el cual $m < 0$ y el sentido de propagación es opuesto al de la rotación. En el caso de modos progrado, una rotación bajo los supuestos mencionados anteriormente provoca un ligero incremento en los períodos, tendiendo a una constante a medida que aumenta la velocidad de rotación. Para modos retrógrados los períodos decrecen a medida que aumenta la fuerza de Coriolis, aunque su evolución cuando se incrementa la velocidad de rotación no es clara.

Cuando existen gradientes químicos en la zona convectiva de la estrella el espaciamiento de los períodos se ve fuertemente afectado [30]. En este caso el espaciamiento de los períodos presenta una componente oscilatoria modelada por el periodo $P_{n,l}$. La periodicidad y amplitud de estas oscilaciones se ven afectadas por la ubicación e intensidad del gradiente respectivamente. Miglio et al. (2008) desarrollaron la teoría y modelos aproximados, además de observaciones. Los autores concluyeron que el espaciamiento entre dos períodos consecutivos puede ser escrito, en forma aproximada, como:

$$\delta P_{n,l} \propto \frac{1}{P_{n,l}} \frac{\Pi_0}{L} \frac{1 - \alpha^2}{\alpha} \frac{1}{\Pi_\mu^{-1}} \cos \left(2\pi \frac{\Pi_0}{\Pi_\mu} n + \pi \frac{\Pi_0}{\Pi_\mu} \phi' + \frac{\pi}{2} \right) \quad (2.12)$$

Cuando ambos fenómenos son combinados sus efectos siguen siendo visibles en el espaciamiento de períodos. En este caso el espaciamiento de períodos no oscila en torno a un valor medio fijo, sino que es variable. El valor medio entre dos mínimos aumenta para modos progrado y decrece para modos retrógrados. Este efecto se puede apreciar en la figura 2.3, que muestra los espaciamientos teóricos de una estrella Gamma Doradus incluyendo efectos de rotación y gradientes químicos. Estos valores fueron simulados por Van Reeth et al. [20] y se encuentran disponibles en CDS [22].

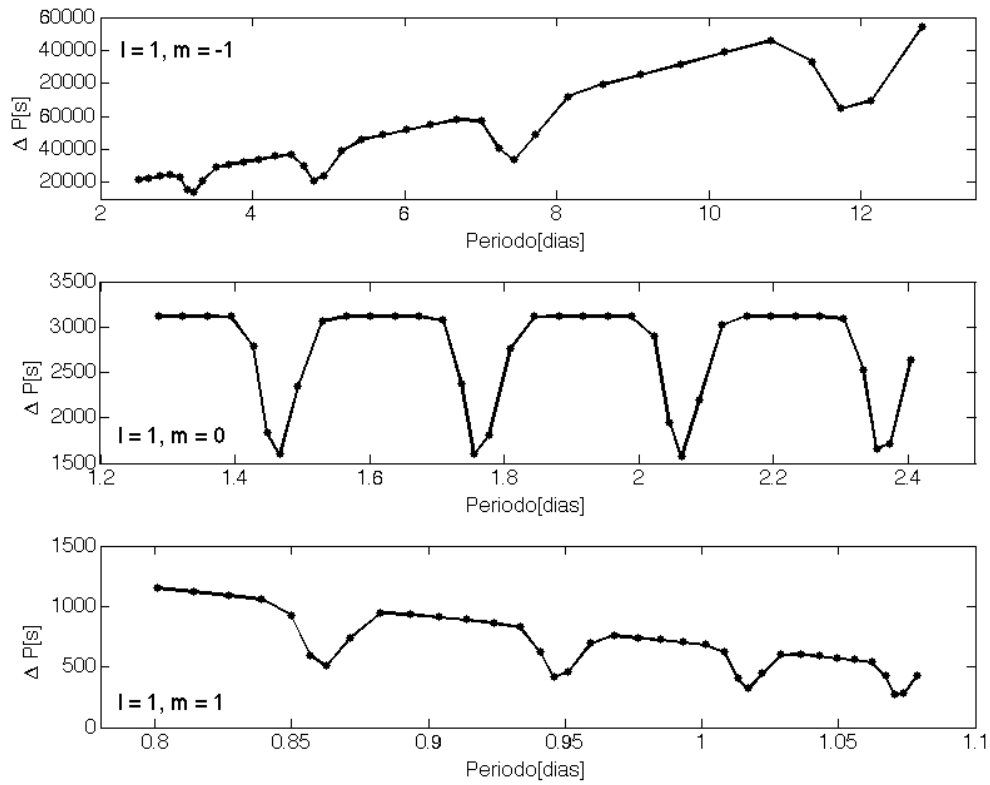


Figura 2.3: Espaciamiento de períodos para simulaciones de Gamma Doradus realizadas por Van Reeth et al. (2015) [20]. Los datos fueron obtenidos de CDS [22]. Las simulaciones incluyen efectos de gradientes químicos y rotación. El efecto del gradiente químico se observa por la componente oscilatoria del espaciamiento. Las tendencias en los modos progrado ($m = 1$) y retrógrado ($m = -1$) son efectos de la rotación.

2.2.4. Métodos de determinación de períodos de pulsación estelar

La disciplina de Asterosismología basa sus estudios en información sobre frecuencias de oscilación y modos de pulsación de las estrellas. Para obtener estos datos se utilizan dos

formas de observación principales: observaciones fotométricas de la variabilidad del flujo de la estrella, y observaciones espectroscópicas de las variaciones de velocidad dadas por los movimientos de elementos en la superficie estelar. Esto es posible debido a que la oscilación estelar causa perturbaciones en el flujo como en la velocidad de una estrella [9].

Para las observaciones fotométricas es fundamental tener una alta precisión con respecto a los tiempos en que se obtienen los datos y en el conteo de electrones. A partir de este conteo se construyen las curvas de luz de luminosidad aparente. Para observaciones en telescopios ubicados en la tierra, es conocida la existencia de períodos espurios al analizar las curvas de luz. Estos períodos se encuentran asociados al hecho que la irregularidad de las observaciones no es aleatoria, sino que producto de condiciones ambientales, tales como las horas de luz en el día. Algunos de estos períodos espurios corresponden al día terrestre, el periodo lunar, o el ciclo anual [31]. Para observaciones en telescopios espaciales, tal como Kepler, estos problemas no ocurren.

En el caso de observaciones espectroscópicas los datos que se obtienen directamente corresponden al flujo correspondiente a cada longitud de onda del espectro. Un *perfil de línea* corresponde a una línea espectral aislada en el espectro de una estrella. La forma de tal línea puede sufrir cambios en el tiempo debido a alteraciones en la velocidad de la superficie de la estrella, temperatura, gravedad o elementos químicos no-uniformemente distribuidos en la superficie, o una combinación de ambos factores. A partir de estos cambios se puede obtener la *medida de velocidad radial baricéntrica*, que representa la componente de velocidad en la línea de visión en ausencia de efectos relativísticos [32], y que se observa como un desplazamiento en el espectro. Usualmente, pero no siempre, las velocidades radiales son suficientes para derivar las frecuencias de pulsación de una estrella.

Tanto en observaciones fotométricas como espectroscópicas es necesario realizar un proceso de identificación de períodos, ya sea en la curva de luz o en la curva de velocidades radiales. Diversos métodos se utilizan, algunos creados especialmente para datos astronómicos, los cuales muchas veces no se encuentran equiespaciados en el tiempo [33]. La discusión de cómo estimar períodos en datos astronómicos computacionalmente fue iniciada por Lafler y Kinman en 1965 [31] quienes propusieron un método de regresión con un criterio basado en mínimos cuadrados. Posteriormente, otras técnicas fueron propuestas, entre ellas el método basado en mínimos cuadrados de Lomb [11] el cual fue extendido posteriormente por Scargle [12], que sigue siendo ampliamente utilizado en el área astronómica. Otras técnicas incluyen el método de transformada de Fourier compensada propuesto por Ferraz-Melo en 1981 [34], el método de largo de *string* de Dworetzky en 1983 [35], análisis de varianza en 1989 por Schwarznberg-Czerny [36], y más recientemente métodos basados en Correntropía por Huijse et al. [37, 38, 39] entre otros. Algunos de estos métodos serán discutidos en la sección 2.3.

Recientemente, Van Reeth et al. (2015) [20] propusieron un método para detectar espaciamientos no uniformes en estrellas Gamma Doradus, combinando información sobre la estrella con sus curvas de luz y aplicándolo a estrellas escogidas del proyecto Kepler. Este método consiste en un conjunto de reglas iterativas en las que se extraen períodos de la curva de luz utilizando Lomb-Scargle y se analiza si son relevantes mediante un análisis de espaciamiento con otros períodos ya aceptados. En simulaciones este método logra distinguir tanto las características asociadas a rotación como aquellas asociadas a gradientes químicos.

2.3. Métodos de análisis de frecuencia

Sea $x(t)$ una señal en tiempo continuo con valores reales, también llamada serie de tiempo, con la propiedad de que su función de autocovarianza $C(\tau)$ (ecuación 2.13) existe y es finita para cada τ . Entonces $C(0)$ (ecuación 2.14) tiene valor finito y es llamada la potencia de la señal $x(t)$ [40].

$$C(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t + \tau)x(t)dt, \quad (2.13)$$

$$C(0) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)^2 dt. \quad (2.14)$$

Wiener [41] estableció la existencia de una función acotada y no decreciente, dependiente de $\lambda = 2\pi f$ ($F(\lambda)$), llamada función de distribución espectral, tal que

$$C(\tau) = \int_{-\infty}^{\infty} e^{i\lambda\tau} F(d\lambda), \quad (2.15)$$

de donde se puede obtener la potencia de la señal para todas las frecuencias menores a λ como $F(\lambda) = F((-\infty, \lambda])$. En la práctica, las series de tiempo son muestreadas en tiempo discreto, por lo que su espectro también puede ser estimado discretamente.

Para transformar una serie de tiempo en tiempo continuo a tiempo discreto se debe definir un muestreo, o forma de escoger los instantes de tiempo discretos. El muestreo de una serie de tiempo puede realizarse de forma regular, con intervalos de tiempo equiespaciados, o de forma irregular. Por ejemplo, al digitalizar sonido, se define una frecuencia de muestreo y se toman datos de forma regular; pero las observaciones astronómicas realizadas desde la superficie terrestre usualmente se encuentran irregularmente muestreadas. Esto último se debe a factores como la imposibilidad de realizar observaciones de día, la disponibilidad de telescopios para realizar observaciones, o la forma en que se eligen los objetivos del telescopio para maximizar las observaciones por noche.

En el caso de muestreos regulares, se debe escoger la distancia temporal entre muestras sucesivas, Δt . Luego se construye una serie de tiempo discreta $X(k)$ tal que,

$$X(k) = x(k\Delta t), k = \dots, -1, 0, 1, \dots \quad (2.16)$$

La discretización de una serie de tiempo no se encuentra exenta de problemas, dado que ciertos armónicos se vuelven indistinguibles entre ellos, por el fenómeno llamado *aliasing*. De esto se desprende el compromiso entre frecuencia de muestreo y resolución espectral. Según el teorema del muestreo, para representar una señal continua con espectro acotado y frecuencia máxima f_{max} se debe escoger una frecuencia de muestreo f_s tal que satisfaga $f_s \geq f_{Ny} = 2f_{max}$, donde f_{Ny} es llamada frecuencia de Nyquist [40].

2.3.1. Análisis espectral de Fourier

Cuando la función de distribución espectral es continua, llamamos a $f(\lambda) = F'(\lambda)$ la función de densidad espectral. En este caso la covarianza corresponde a la Transformada de Fourier inversa de $f(t)$ (ecuación 2.17), o equivalentemente, la densidad espectral de potencia se define como la Transformada de Fourier de la covarianza de la señal (ecuación 2.18).

$$C(\tau) = \int_{-\infty}^{\infty} e^{i\lambda\tau} f(\lambda) d\lambda \quad (2.17)$$

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C(\tau) e^{-i\lambda\tau} d\tau \quad (2.18)$$

El espectro de una señal discreta también puede ser obtenido a partir de la transformada de Fourier de su covarianza. Utilizando una aproximación determinada por la cantidad de muestras disponibles para el cómputo del espectro de Fourier, se define el periodograma como $I_N(\lambda)$:

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{i\lambda k} C(k) \quad (2.19)$$

$$I_N(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{i\lambda k} \tilde{C}(k). \quad (2.20)$$

donde \tilde{C} corresponde a la estimación utilizada de la covarianza, por ejemplo la de la ecuación 2.21, aunque otros estimadores pueden ser aplicados.

$$\tilde{C}(k) = \begin{cases} \frac{1}{N} \sum_{t=1}^{N-|k|} X(t+|k|)X(t) & \text{si } |k| \leq N-1, \\ 0 & \text{si } |k| > N-1 \end{cases} \quad (2.21)$$

En términos computacionales, la forma más eficiente de estimar el espectro de frecuencia de una señal muestreada en tiempo discreto es la Transformada Rápida de Fourier (FFT) [42], que presenta una complejidad computacional de $\mathcal{O}(N \log_2 N)$ para el cómputo del espectro de una señal con N puntos.

2.3.2. Periodograma de Lomb-Scargle

Propuesto originalmente por Lomb [11] y posteriormente extendido por Scargle [12], el periodograma de Lomb-Scargle es un método basado en mínimos cuadrados para obtener una representación en frecuencia, particularmente cuando se tienen señales muestreadas de manera irregular, como es el caso de los datos astronómicos.

Dada una serie de tiempo X , correspondiente a la medición de una señal X_s en un conjunto de instantes de tiempo t_i , $X(t_i)$, $i = 1, 2, \dots, N_0$, tal que las muestras pueden contener ruido aditivo $\sigma(t_i)$,

$$X_i = X(t_i) = X_s(t_i) + \sigma(t_i), \quad (2.22)$$

entonces la potencia asociada a cada frecuencia $\omega = 2\pi f$ se calcula como

$$P_X(\omega) = \frac{1}{2} \left[\frac{[\sum_i X_i \cos(\omega(t_i - \tau))]^2}{\sum_i \cos(\omega(t_i - \tau))} + \frac{[\sum_i X_i \sin(\omega(t_i - \tau))]^2}{\sum_i \sin(\omega(t_i - \tau))} \right] \quad (2.23)$$

donde el valor de τ está dado por

$$\tan(2\omega\tau) = \frac{\sum_i \sin(2\omega t_i)}{\sum_i \cos(2\omega t_i)}. \quad (2.24)$$

Este tipo de análisis de frecuencia presenta ciertas ventajas en comparación con el periodograma clásico de Fourier. En primera instancia, puede ser usado para series de tiempo de muestreo irregular. Adicionalmente, en series de tiempo muestreadas irregularmente (e idealmente de forma aleatoria) un gran número de frecuencias puede ser analizado, debido a que en estos casos la frecuencia de Nyquist equivalente puede llegar a ser mucho más alta que en el caso de un muestreo regular [43].

Adicionalmente, el periodograma de Lomb-Scargle permite establecer probabilidades de falsa alarma para los valores de potencia asociados a frecuencias o períodos. De esta manera, si se define $Z = P_X(\omega)$ es posible estudiar la distribución de probabilidades, donde la función de probabilidad acumulada está dada por

$$F_Z(z) = \mathbb{P}(Z < z) = \int_0^z p_Z(z') dz' = 1 - \exp(-z). \quad (2.25)$$

Por otro lado, si se considera un valor máximo en el espectro, donde $Z = \max_n P(\omega_n)$ es el máximo sobre un conjunto de frecuencias ω_n entonces se obtiene que

$$\mathbb{P}(Z > z) = 1 - [1 - \exp(-z)]^N. \quad (2.26)$$

A partir de esto se puede obtener el valor mínimo que debe tener un elemento del espectro para que su probabilidad de falsa alarma sea p_{fa} , el cual es deseable que sea un valor muy pequeño, donde N es la cantidad de frecuencias independientes analizadas:

$$z_0 = -\ln [1 - (1 - p_{fa})^{1/N}]. \quad (2.27)$$

Periodograma de Lomb-Scargle Generalizado

Pese a sus ventajas, el periodograma de Lomb-Scargle presenta algunos problemas. El año 2009, Zechmeister y Kürster propusieron el periodograma de Lomb-Scargle generalizado para resolver dos defectos del método original: la falta de consideración de los errores en las mediciones, y el supuesto de que las medias de los datos y de las señales sinusoidales a ser ajustadas son iguales [44].

Para ello se consideran funciones sinusoidales completas, incluyendo un sesgo: $y(t) = a \cos(\omega t) + b \sin(\omega t) + c$, y se busca minimizar el siguiente funcional:

$$\chi = \sum_{i=1}^N \frac{[X_i - y(t_i)]^2}{\sigma_i^2} = W \sum w_i [X_i - y(t_i)]^2, \quad (2.28)$$

donde los pesos w_i son dependientes de los errores en las mediciones, σ_i , de la siguiente manera:

$$w_i = \frac{1}{W} \frac{1}{\sigma_i} \quad \left(W = \sum \frac{1}{\sigma_i^2} \quad \sum w_i = 1 \right). \quad (2.29)$$

La potencia asociada a cada frecuencia la siguiente:

$$p(\omega) = \frac{\chi_0^2 - \chi(\omega)^2}{\chi_0^2}, \quad (2.30)$$

donde χ_0 es el funcional χ evaluado en la media ponderada $\bar{X} = \sum w_i X_i$.

2.3.3. Otros métodos de análisis de frecuencia

Tal como fue comentado en la sección 2.2.4 numerosos métodos de estimación de espectro han sido propuestos, algunos específicamente para lidiar con las características de los datos astronómicos. La versión generalizada de Lomb-Scargle es similar a la solución propuesta por Ferraz-Mello [34], aunque realizando una conexión con Lomb-Scargle en vez de con la transformada de Fourier Discreta. Además de Lomb-Scargle y su versión generalizada, los métodos de análisis de varianza [36] y largo de cuerda [35] son ejemplos de métodos que han sido utilizados para la determinación de períodos de oscilación.

El método de largo de cuerda consiste en minimizar la suma de los largos de segmentos que unen puntos sucesivos en un diagrama de fase, que puede ser utilizado cuando existen muy pocos datos en un lapso largo de tiempo. Por otro lado, el método de análisis de varianza propone utilizar significancia estadística en el análisis de períodos, dado el ruido inherente que tienen los datos.

Otros métodos también han sido propuestos y utilizados en series de tiempo astronómicas, como aquellos basados en Teoría de la Información para el Aprendizaje: *Correntropy Kernelized Periodogram* o CKP, basado en Correntropía, una generalización de la Correlación. [37, 38]; y espectros basados en descomposiciones matriciales no-negativas de Correntropía [39].

2.4. Autómata celular

Un Autómata celular es una colección de celdas coloreadas en una grilla pre-definida, donde el estado de activación de cada celda evoluciona en pasos de tiempo discretos, según reglas determinadas de acuerdo al estado de su vecindario. En el caso más básico se tienen reglas determinísticas, por ejemplo, que el estado de activación corresponde a la moda de los estados de las celdas del vecindario.

Hay cuatro tipos de comportamientos cualitativos que se pueden obtener, dada una inicialización aleatoria y dependiendo de las reglas de evolución [13]:

1. Tiende a un estado espacialmente homogéneo.
2. Lleva a una secuencia de estructuras simples o periódicas.
3. Exhibe un comportamiento caótico aperiódico.
4. Lleva a estructuras localizadas complicadas, que en algunos casos se propagan.

Estos comportamientos han sido asociados a ciertos tipos de sistemas dinámicos. Por ejemplo, los primeros tres son análogos a puntos fijos, ciclos límite y atractores caóticos. Las distintas clases presentan comportamientos diferentes en cuanto a los patrones observables en la evolución y los efectos que tienen pequeños cambios en la configuración inicial.

El Autómata Celular (determinístico) más famoso es el llamado Juego de la Vida, o *Game of Life* en inglés, inventado por Conway en 1970. En este ejemplo se tiene una grilla bidimensional con un vecindario de Moore, donde las ocho celdas que rodean una posición corresponden a su vecindario. El juego define únicamente dos estados: vivo y no vivo. La regla de actualización señala que un individuo puede mantenerse vivo si y solo si dos o tres de sus vecinos también están vivos. Si hay menos, entonces la célula muere por aislamiento, y si hay más, muere por sobrepoblación. Una célula que no se encuentra viva cambia de estado sólo si tiene exactamente tres vecinos vivos. Diferentes inicializaciones aleatorias del Juego de la Vida pueden llevar a distintas configuraciones, tales como estados fijos, oscilatorios y "planeadores", los cuales son configuraciones finitas que se trasladan [45].

El Autómata Celular Probabilístico mantiene la misma idea, cambiando únicamente la forma de actualización de los estados. En este caso se introducen reglas probabilísticas para la actualización de los estados.

2.5. Percolación

Es una teoría fundada por Broadbent y Hammersley en 1957 [46]. Originalmente su objetivo era estudiar procesos físicos aleatorios, tales como el flujo de un fluido a través de un medio poroso desordenado, mediante transiciones de fase. Una de las interpretaciones posibles es que permite el estudio de grafos aleatorios, que pueden o no ser orientados y formar grillas. Para crear grafos aleatorios se seleccionan vértices o aristas con probabilidad p en un grafo con estructura pre-definida. Esta probabilidad señala si los espacios del grafo están ocupados o no. La teoría de Percolación trata del estudio de los *clusters* y propiedades de estos grafos aleatorios. Existen dos formas de crear los grafos: por vértices o aristas [47].

Se dice que un grafo se encuentra percolado cuando se puede construir un camino a través del grafo con los espacios ocupados o activos. La probabilidad crítica es aquella en la cual el grafo comienza a percolar. De manera más formal, es la probabilidad en que se puede formar un *cluster* de tamaño infinito en una grilla infinita. Los aspectos relacionados a la percolación cercana a la probabilidad crítica son llamados fenómenos críticos [48].

Bootstrap Percolation es un ejemplo particular de autómatas celulares, que se define como la propagación de un estado de activación [49]. En este caso se comienza con una inicialización aleatoria, y la única evolución posible es desde un estado no-activo a uno activo. No ocurren transiciones en el sentido opuesto. Cuando un lugar se encuentra activo, éste no cambiará su estado. Sin embargo, un sitio inactivo puede cambiar de estado si al menos ℓ de sus vecinos ya se encuentran activos. Este comportamiento se ilustra en la figura 2.4, donde dada la condición inicial mostrada en la grilla superior izquierda, el estado activo se propaga por la grilla en sucesivas iteraciones, con $\ell = 2$.

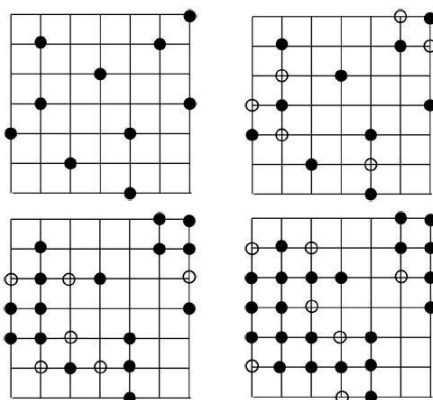


Figura 2.4: Ejemplo de Bootstrap Percolation, donde estados no-activos se transforman a activos cuando al menos dos de sus vecinos ya se encuentran en tal estado. Imagen obtenida de Scholarpedia [14].

2.6. Conjuntos K de Freeman

Son una jerarquía de modelos basados en la dinámica de las poblaciones de neuronas a nivel intermedio, las cuales tienen un tamaño de aproximadamente 10^4 neuronas, que presentan del orden de 10^8 conexiones sinápticas. La topología de las conexiones es modelada por redes de poblaciones de neuronas con actividad inhibitoria o excitatoria. Las dinámicas son aproximadas por linealizaciones por partes de ecuaciones diferenciales ordinarias no lineales. [16, 50]. Diferentes dinámicas son representadas por distintos conjuntos [16, 51], los cuales pueden ser observados en la figura 2.5. Cada conjunto se encuentra inspirado en diferentes estructuras cerebrales, que son parte de una jerarquía de estructuras cada vez más complejas.

La neurona, unidad básica de las dinámicas que se quieren representar, se encuentra formada por diversas estructuras, entre ellas las dendritas y los axones. Las dendritas reciben múltiples estímulos provenientes de otras neuronas, los combinan y posteriormente entregan el resultado al segmento inicial del axón, el que transmite este resultado a otras partes del sistema nervioso. Las dendritas reciben estímulos en forma de pulsos, los cuales transforman a una corriente con variaciones continuas, mientras que los axones reciben esta señal continua transformándola en trenes de pulsos. Estos dos modos corresponden a las variables de estado de los Conjuntos K de Freeman.

Para poder construir los diferentes conjuntos se deben realizar ciertas simplificaciones. En el cerebro, las neuronas se encuentran distribuidas en un espacio tridimensional, aunque para el modelamiento topológico cada conjunto es concebido como una distribución bidimensional de neuronas. Adicionalmente, no se asigna un valor para la actividad de cada neurona por separado, debido a que el estado de una neurona está íntimamente ligado con los de su vecindad. Luego, en cada tiempo t y punto en la superficie (x_1, y_1) existe un estado local de activación $o(t, x_1, y_1)$. Cualquier grupo de mediciones sobre la manifestación de tal estado de activación en una colección de puntos del conjunto K0 permite describir la distribución de actividad, ya sea en el modo de pulso $p(t, \vec{x})$ o en el modo de onda $v(t, \vec{x})$.

El objetivo es simular los patrones espacio temporales de la corteza cerebral. Análisis experimentales han demostrado que la ecuación que describe estos patrones puede ser separada en tres partes: una función de transferencia linealmente dependiente del tiempo $F(t)$; una función dependiente de la amplitud de manera no lineal $G(v)$; y una función dependiente del espacio $H(x)$,

$$v(t, \vec{x}) = F(t)G(v)H(\vec{x}) + I(t). \quad (2.31)$$

En la expresión anterior, $I(t)$ corresponde a un estímulo, usualmente modelado como un impulso. Cuando se considera un único punto en el espacio, entonces $H(x) = 1$.

Los conjuntos alcanzables más básicos son los siguientes:

- **Conjunto K0**

Representa un arreglo de entre 10^3 a 10^8 neuronas con características inhibitorias o excitatorias, pero que no tienen inter-conexiones funcionales, sino que conexiones co-

munes de entradas y salidas. La fuente de la entrada puede ser un estímulo sensorial, tal como un olor, luz, o alguna señal proveniente de un conjunto distinto de neuronas. Los receptores olfativos son un ejemplo de conjuntos K0. En el cerebelo, las células de Golgi exhiben este comportamiento.

La dinámica de la respuesta sumada de las dendritas puede ser modelada mediante la siguiente Ecuación Diferencial Ordinaria, donde $v(t)$ denota la magnitud de la actividad neuronal en modo de onda,

$$\frac{d^2v(t)}{dt^2} + (a + b)\frac{dv(t)}{dt} + ab = k_{jk}\delta(t). \quad (2.32)$$

De esto se obtiene que la función de evolución temporal en el dominio de Laplace es:

$$F(s) = k_{jk} \frac{ab}{(s + a)(s + b)}. \quad (2.33)$$

Un conjunto K0 que modela dinámicas de lazo abierto en las poblaciones neuronales sólo tiene un atractor de punto cero, al cuál regresa luego de la perturbación.

- **Conjunto KI**

Corresponde a poblaciones con excitación o inhibición mutua. Se forman debido a la interconexión de conjuntos K0 excitatorios o inhibitorios mediante un lazo de retroalimentación positiva. Un ejemplo de este conjunto son las células periglomerulares del sistema olfativo, las que pueden causar inhibición en neuronas vecinas.

La dinámica del lazo cerrado, usando la ecuación 2.33 está dada por:

$$V(s) = \frac{F(s)}{1 - k_{ee}F^2(s)}. \quad (2.34)$$

La dinámica de los conjuntos KI se describe como convergencia a un punto fijo distinto de cero.

- **Conjunto KII**

Corresponde a la interacción entre poblaciones excitatorias e inhibitorias densamente conectadas. En este tipo de conjuntos, cada neurona, ya sea excitatoria o inhibitoria, presenta conexiones a otras neuronas tanto excitatorias como inhibitorias. Un ejemplo de este conjunto es el bulbo olfatorio. El caso más simple resulta de la interacción de un set K0 excitatorio con un set KI inhibitorio cuando todas las ganancias son iguales, en el cual se obtiene la siguiente dinámica,

$$V(s) = \frac{F(s)}{[1 + k_{ei}k_{ie}F^2(s)]}. \quad (2.35)$$

Transiciones desde un atractor puntual hacia un ciclo límite son posibles para un nivel adecuado de retroalimentación. En el dominio del tiempo, esto significa que comportamientos oscilatorios pueden ser observados en un punto determinado del espacio.

- **Conjunto KIII**

Éstos se encuentran formados por múltiples conjuntos KII que interactúan. Generan oscilaciones aperiódicas o caóticas mediante la combinación de retroalimentación positiva y negativa de distintos sets KII con frecuencias inconmensurables. Pueden representar patrones de amplitud modulada (AM) espacio-temporales.

- **Conjuntos de mayor complejidad**

Conjuntos KIV, KV o superiores pueden ser creados mediante la interacción de conjuntos de orden inferior.

2.7. Neuropercolación

Neuropercolación es una familia de modelos estocásticos basados en la teoría matemática de Autómata Celular Probabilístico en grillas y grafos aleatorios. Se encuentra motivado por las propiedades dinámicas y estructurales de las poblaciones neuronales. La existencia de transiciones de fase ha sido demostrada en espacios de estado continuos y discretos. Neuropercolación extiende el concepto de transiciones de fase a grandes poblaciones de células nerviosas que interactúan entre sí [14, 15]. Esta familia de modelos presenta dinámicas inspiradas por los Conjuntos K propuestos por Walter Freeman en 1975 [16].

Freeman describió diez elementos fundamentales que permiten comprender las dinámicas de las poblaciones neuronales [53]. Cinco de estos principios son posibles de representar utilizando Neuropercolación:

1. La transición de estado de una población excitatoria desde un atractor puntual con actividad cero hacia un atractor distinto de cero con actividad de estado estable mediante retroalimentación positiva.
2. Aparición de oscilación a través de retroalimentación negativa entre poblaciones neuronales excitatorias e inhibitorias.
3. Transiciones de estado desde un atractor puntual hacia un atractor de ciclo límite, que regula oscilaciones de estado estable de una población cortical mixta excitatoria-inhibitoria.
4. Génesis de caos como actividad de segundo plano, mediante la combinación de retroalimentación positiva y negativa entre tres o más poblaciones mixtas excitatorias-inhibitorias.
5. Ondas distribuidas de actividad caótica que llevan a un patrón espacial de amplitud modulada construida en base a las amplitudes locales de la onda.

Neuropercolación describe estos efectos en modelos discretos, en comparación a los modelos K de Freeman que los definen en tiempo continuo.

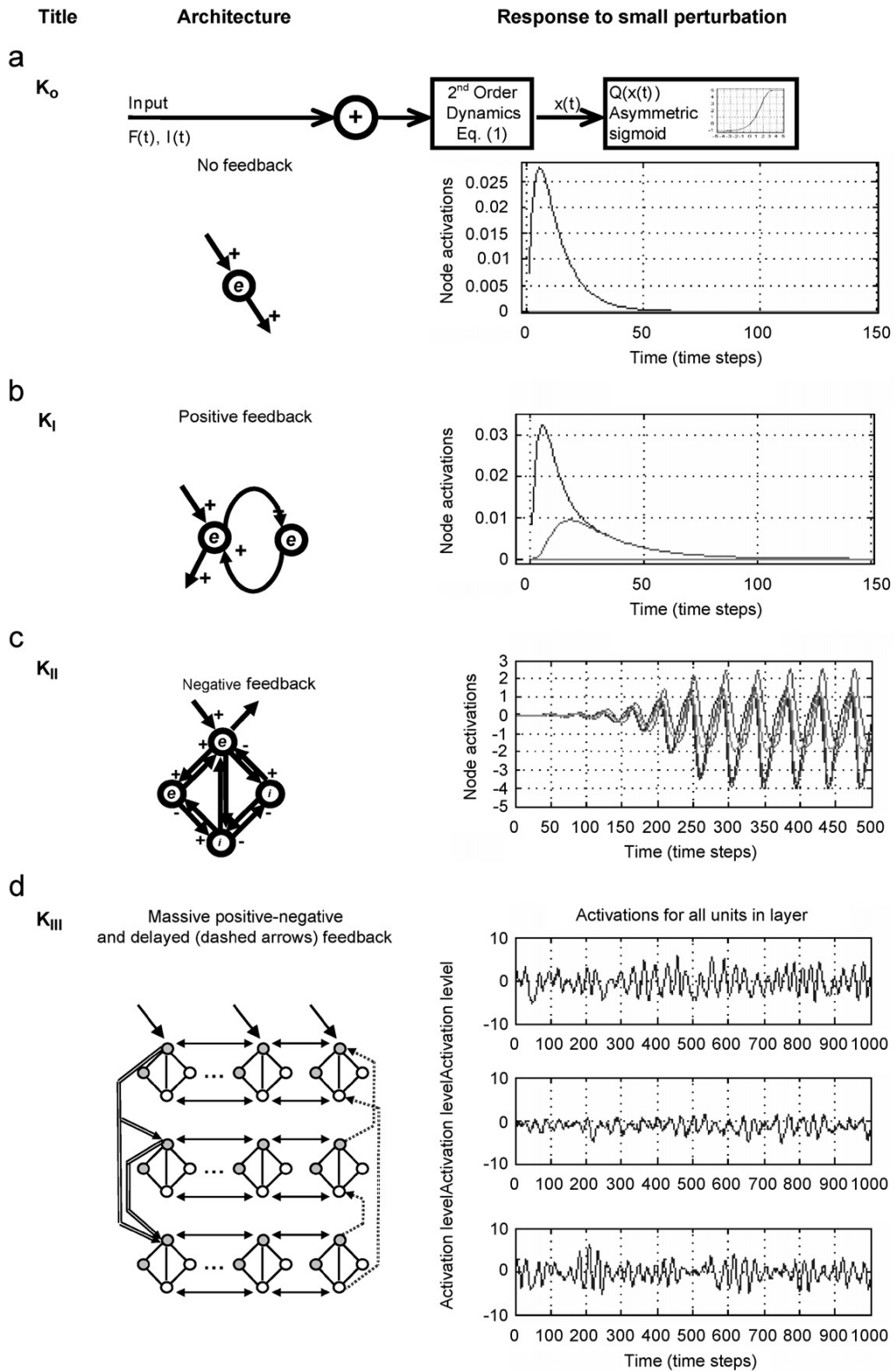


Figura 2.5: Conjuntos K de Freeman (K_0 a K_{III}) según la respuesta neuronal que representan. Obtenido de Beliaev y Kozma, 2007 [52].

Bootstrap Percolation presenta limitaciones que el modelo de Neuropercolación pretende relajar, de manera de representar las propiedades de las poblaciones neuronales y obtener un

modelo más general. En primer lugar, luego de una inicialización aleatoria, *Bootstrap Percolation* es un proceso determinístico. Adicionalmente, la única transición de estado posible es desde un estado no-activo hacia uno activo. Las siguientes propiedades son incorporadas en el modelo de Neuropercolación [14]:

1. **Interacción con el ruido:** Dado que las dinámicas naturales del cerebro son no-determinísticas, Neuropercolación incorpora la aleatoriedad como parte del modelo, en las reglas de evolución.
2. **Efecto de axones largos:** En una configuración típica de Autómata Celular Probabilístico las conexiones que forman vecindades son cortas, relacionadas a los vecinos topográficos. Al re-cablear algunos vértices de la grilla para permitir conexiones entre neuronas a mayores distancias, y no sólo con sus vecinos inmediatos, se pueden obtener transiciones de fase. De esta manera, dependiendo de la proporción de vecinos lejanos se pueden presentar efectos de ciclo límite sostenidos, así como oscilaciones no periódicas, similares a aquellas que se pueden obtener de ecuaciones diferenciales acopladas.
3. **Inhibición:** Adicionalmente a ser un efecto presente ampliamente en estructuras cerebrales, el efecto de inhibición contribuye a la presencia de dinámicas oscilatorias y controla la estabilidad. Adicionalmente permite volver a estados de inactividad. Efectos tanto excitatorios como inhibitorios se encuentran presentes en los modelos de Neuropercolación.

Neuropercolación utiliza la jerarquía de poblaciones interactuando en redes desarrolladas en los modelos K de Freeman, pero reemplaza las ecuaciones diferenciales ordinarias con distribuciones de probabilidad de redes aleatorias que evolucionan en el tiempo. Estos equivalentes pueden observarse en la figura 2.6. Un conjunto KI es representado por un único grafo con propiedades excitatorias o inhibitorias (Figura 2.6(a)). Al interconectar dos subgrafos, uno excitatorio y el otro inhibitorio, se obtiene un conjunto KII (Figura 2.6(b)). La interconexión de múltiples conjuntos KII en forma de grillas permite la obtención de un conjunto KIII (Figura 2.6(c)). La topología utilizada para construir un conjunto KIII no es única, pudiendo variar el número de osciladores que lo componen, la cantidad de interconexiones entre conjuntos KII y el orden de los elementos de la grilla. La teoría de Neuropercolación provee un acercamiento matemático para describir transiciones de fase y fenómenos críticos en redes de gran escala [51].

En Autómatas Celulares, los estados de un punto en la grilla son activos o inactivos. A cada instante se actualiza el estado de los puntos mediante reglas determinísticas o probabilísticas dependientes del vecindario de cada punto. En los modelos *Mean Fields*, en vez de considerar el número de nodos activos en una vecindad Γ , las activaciones de $|\Gamma|$ nodos de la grilla seleccionados aleatoriamente son utilizados en la regla de activación (con reemplazo). En este modelo, no hay dependencia de la topología de la grilla [14]. Según lo anterior, al considerar todas las opciones de conexiones de vecinos en las grillas, se pueden reemplazar las ecuaciones diferenciales por reglas probabilísticas.

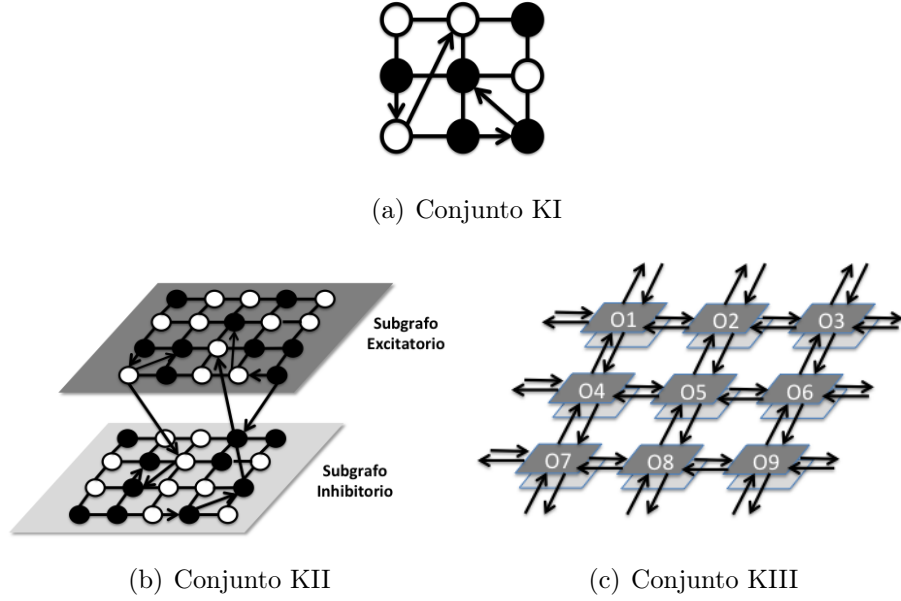


Figura 2.6: Conjuntos K de Freeman utilizando Neuropercolación. Un conjunto KI corresponde a un arreglo de neuronas que interactúan entre sí de manera inhibitoria o excitatoria (2.6(a)). Un conjunto KII (2.6(b)) corresponde a dos subgrafos KI, uno excitatorio y otro inhibitorio, que interactúan entre sí. Los conjuntos KIII (2.6(c)), que son aquellos que pueden representar actividad caótica y cambios de fase, se representan mediante arreglos de osciladores (o conjuntos KII) interconectados. En los conjuntos KI y KII, los nodos de color negro simbolizan el estado activo, mientras que aquellos nodos blancos se encuentran inactivos.

2.7.1. Ecuaciones del modelo *Mean Fields*

En una grilla, si cada vértice modela una neurona, entonces las dendritas y el axón son descritas por las aristas. Se considera un grafo $G(\mathcal{V}, E)$ con vértices \mathcal{V} y aristas E . Un grafo de campo-medio (*mean-field*) $G(t, \mathcal{V}, E(t))$ que evoluciona en el tiempo se define de la siguiente manera: Sean $\mathcal{N}(t, \mathbf{v}_x)$ los vecinos del vértice $\mathbf{v}_x \in \mathcal{V}$, que son seleccionados de manera independiente con repetición desde el conjunto de todos los vértices. Se pueden definir diversos modelos de campo-medio seleccionando el tamaño de vecindario d , que también es llamado orden del modelo de campo-medio. El conjunto de aristas que conectan los vecindarios se define como $E(t)$. Dado un vecindario, el estado de activación de cada neurona, o vértice en el grafo, evoluciona mediante reglas de mayoría probabilística, lo que significa que un vértice \mathbf{v}_x sigue el estado de activación de la mayoría de sus vecinos con una alta probabilidad ω .

En las aproximaciones de campo-medio es suficiente conocer el estado esperado de actividad sobre la grilla, a_t , para calcular el estado de actividad en el instante siguiente a_{t+1} . De esta manera, el estado esperado de cada vértice \mathbf{v}_x en un instante t es la probabilidad de que esté en estado activo.

Dado el contexto previo, se pueden obtener expresiones recursivas para los estados de activación en los diferentes conjuntos K de Neuropercolación [54].

Nivel KI

En este nivel consideramos un único grafo $G(t, \mathcal{V}, E(t))$ con tamaño de vecindad d . Cada vértice \mathbf{v}_x presenta un estado de activación discreto (1 o 0), que se modifica de acuerdo a la regla de la mayoría. Existe una probabilidad ω de que siga el estado de la mayoría de sus vecinos, y una probabilidad $1 - \omega$ de que haga lo contrario. De esta manera, para un vecindario $\mathcal{N}(t, \mathbf{v}_x)$ hay $\binom{d}{i}$ maneras de elegir i vecinos en el estado 0, siendo el estado esperado de activación de los posibles vecindarios $a_t^{d-i}(1 - a_t)^i$, donde i es la cantidad de vecinos inactivos y a_t es el estado promedio de activación. Con esta información, el estado esperado de activación del grafo puede ser obtenido recursivamente mediante la fórmula siguiente,

$$a_{t+1} = \sum_{i=0}^d \binom{d}{i} a_t^{d-i} (1 - a_t)^i f_i(\omega). \quad (2.36)$$

$$f_i(\omega) \triangleq \begin{cases} \omega & \text{si } i < d/2 \\ 0 & \text{si } i = d/2 \\ 1 - \omega & \text{si } i > d/2 \end{cases} \quad (2.37)$$

Esto se puede reescribir definiendo lo siguiente:

$$a_{t+1} = \omega N_\omega(d, a_t) + N_0(d, a_t) + N_m(d, a_t) \quad (2.38)$$

$$N_\omega(d, a_t) \triangleq \sum_{i=0}^{\lfloor \frac{d}{2} \rfloor} \binom{d}{i} (a_t^{d-i} (1 - a_t)^i - (1 - a_t)^{d-i} a_t^i) \quad (2.39)$$

$$N_0(d, a_t) \triangleq \sum_{i=0}^{\lfloor \frac{d}{2} \rfloor} \binom{d}{i} (1 - a_t)^{d-i} a_t^i \quad (2.40)$$

$$N_m(d, a_t) \triangleq \begin{cases} 0 & \text{si } d \text{ es impar} \\ \frac{1}{2} \binom{d}{\frac{d}{2}} (a_t(1 - a_t))^{\frac{d}{2}} & \text{si } d \text{ es par} \end{cases} \quad (2.41)$$

Nivel KII

El Nivel KII se representa como dos grafos de nivel KI interconectados, uno cumpliendo la labor inhibitoria y el otro la excitatoria. Sean dos grafos $G_0, G_1 \in KI$, donde los vértices de G_0 excitan los vértices de G_1 , y los vértices de G_1 inhiben a los de G_0 . Si $\mathbf{v}_x \in G_0$ excita a $\mathbf{v}_{x'} \in G_1$ con probabilidad p^+ mientras que $\mathbf{v}_{x'}$ inhibe a \mathbf{v}_x con probabilidad p^- , entonces los estados esperados de activación de \mathbf{v}_x y de $\mathbf{v}_{x'}$ son, respectivamente, los siguientes:

$$\alpha_{t,0} = (1 - p^-)a_{t,0} + p^-(1 - a_{t,1}) \quad (2.42)$$

$$\alpha_{t,1} = (1 - p^+)a_{t,1} + p^+a_{t,0}. \quad (2.43)$$

Si los grafos de campo-medio G_0 y G_1 son del mismo orden, entonces el estado de activación esperado de los grafos completos es el promedio de los estados esperados de los vértices $\mathbf{v}_x \in G_0$ y $\mathbf{v}_{x'} \in G_1$, los cuáles se calculan de la misma manera que en el caso de los niveles KI, obteniendo lo que sigue:

$$a_{t+1,0} = \omega N_\omega(d, \alpha_{t,0}) + N_0(d, \alpha_{t,0}) + N_m(d, \alpha_{t,0}) \quad (2.44)$$

$$a_{t+1,1} = \omega N_\omega(d, \alpha_{t,1}) + N_0(d, \alpha_{t,1}) + N_m(d, \alpha_{t,1}). \quad (2.45)$$

Bajo esta configuración hay tres dinámicas posibles para los estados de activación promedio de ambos grafos, donde ω_0 y ω_1 son probabilidades críticas que delimitan tales dinámicas:

- i. Convergencia a un atractor donde $a_{t,0} = a_{t,1} = \frac{1}{2}$, cuando $\omega < \omega_0$.
- ii. Convergencia a un ciclo límite, donde $a_{t,0}$ y $a_{t,1}$ oscilan alrededor de $\frac{1}{2}$, si $\omega_0 < \omega < \omega_1$.
- iii. Convergencia a un atractor distinto de $\frac{1}{2}$, cuando $\omega > \omega_1$.

Nivel KIII

Dos o más osciladores KII producen el grafo de un conjunto KIII. De esta manera, se conectan M osciladores en un arreglo, resultando el grafo $G(t, \mathcal{V}, E, 2M) = \bigcup_{i=0}^{2M-1} (G_i \cup G_{i+1})$, donde los índices pares corresponden a los grafos con conexiones excitatorias, mientras que los índices impares señalan a aquellos con conexiones inhibitorias. Al complejizar los modelos, los osciladores interactúan entre ellos, por lo que se definen tres tipos de conexiones:

- Vértices $\mathbf{v}_x \in G_i$ son inhibidos por vértices $\mathbf{v}_{x'} \in G_{i+1}$ con probabilidad $p_{i/2}^-$
- Vértices $\mathbf{v}_x \in G_{i+1}$ son excitados por vértices en G_i con probabilidad $p_{i/2}^+$
- Vértices en G_i son excitados por vértices en G_j con probabilidad $c_{j,i}$, donde j es seleccionado desde un conjunto de índices $\mathcal{N}_i \subset 0, 2, \dots, 2M - 2$ que representa el vecindario de G_i en el arreglo.

De esta manera, a_t corresponde al estado esperado sobre todos los osciladores del grafo $G(t, \mathcal{V}, E, 2M)$, mientras que $a_{t,l}$ es el estado esperado del oscilador l . Los estados esperados de los vértices $\mathbf{v}_x \in G_i$ y $\mathbf{v}_{x'} \in G_{i+1}$ para índices i pares se denotan $\alpha_{t,i}$ y $\alpha_{t,i+1}$ respectivamente, y se calculan como sigue:

$$\alpha_{t,i} = \left(1 - p_{i/2}^- - \sum_j c_{j,i} \right) a_{t,i} + p_{i/2}^-(1 - a_{t,i+1}) + \sum_j c_{j,i} a_{t,j} \quad (2.46)$$

$$\alpha_{t,i+1} = \left(1 - p_{i/2}^+ \right) a_{t,i+1} + p_{i/2}^+ a_{t,i}. \quad (2.47)$$

Si los grafos G_i que componen $G(t, \mathcal{V}, E, 2M)$ son del mismo orden d , entonces los estados esperados de éstos son:

$$a_{t+1} = \frac{1}{2M} \sum_{i=0}^{2M-1} a_{t+1,i} \quad (2.48)$$

$$a_{t+1,i} = \omega_i N_\omega(d, \alpha_{t,i}) + N_0(d, \alpha_{t,i}) + N_m(d, \alpha_{t,i}) \quad (2.49)$$

2.8. Optimización por enjambre de partículas

Optimización por enjambre de partículas, o PSO por sus siglas en inglés, es un método de optimización basado en poblaciones que fue propuesto por Kennedy y Eberhart [18, 19]. En este método, cada partícula realiza una búsqueda en el espacio de variables con el objetivo de encontrar una solución óptima. La posición de una partícula cambia de acuerdo a reglas preestablecidas que incluyen una componente social y una personal. El algoritmo PSO original modifica la velocidad de cada partícula según

$$v_i^{t+1} = v_i^t + \varphi_1 U_1^t (b_i^t - x_i^t) + \varphi_2 U_2^t (l_i^t - x_i^t) \quad (2.50)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (2.51)$$

donde φ_1 y φ_2 son coeficientes de aceleración, U_1^t y U_2^t son coeficientes generados por distribuciones uniformes aleatorias, l_i es la mejor posición encontrada en una vecindad de la partícula i , y b_i es la mejor posición personal de la partícula. Una de las variantes de PSO corresponde a aquella con coeficiente de constricción χ que pondera la ecuación 2.50 para la velocidad [55], como sigue:

$$v_i^{t+1} = \chi^t [v_i^t + \varphi_1 U_1^t (b_i^t - x_i^t) + \varphi_2 U_2^t (l_i^t - x_i^t)]. \quad (2.52)$$

El coeficiente de constricción provoca que el movimiento de cada partícula vaya disminuyendo con el tiempo. De esta manera, la mayor exploración del espacio ocurre en las primeras iteraciones, para posteriormente buscar los óptimos en espacios más reducidos.

PSO resulta útil cuando es infactible realizar un método exacto de optimización, por ejemplo gradiente descendente. Adicionalmente, puede ser utilizado para problemas de alta dimensionalidad. Pese a ello tiene problemas como que puede entregar mínimos locales y que tiene una aleatoriedad intrínseca debido a la forma en que evoluciona la población. Por este último punto es necesario repetir la optimización múltiples veces para poder hacer un análisis de los resultados. En el caso particular planteado en este trabajo, PSO es el método de optimización escogido debido a que el modelo no es una función sino un caso particular de autómatas celulares probabilísticos, donde la obtención del gradiente puede llegar a ser compleja.

3 Metodología

Tradicionalmente los modelos de Neuropercolación han sido utilizados con topologías de grillas. En este trabajo se estudia, además de éstas, una nueva topología propuesta con forma esférica, inspirada en la dinámica estelar. Dada una serie de tiempo que se quiere modelar mediante Neuropercolación, se debe definir la estructura del modelo a utilizar: ya sea un modelo plano o esférico. Una vez definida la estructura, se debe determinar el tamaño que tendrá ésta, lo que definirá la cantidad de osciladores que tenga el modelo y cómo éstos se encuentran interconectados. En el caso de los modelos planos se debe definir el tamaño de la grilla, mientras que para los modelos esféricos se debe especificar la cantidad de capas. Independiente de la estructura, dada una cantidad N de osciladores, con K conexiones orientadas desde y hacia otros osciladores, existen $N \cdot (3 + K)$ parámetros a determinar. Estos serían N valores de probabilidad de seguir a la mayoría (ω_i), probabilidad de inhibición p_i y tamaño de la vecindad d_i , uno para cada oscilador; además de $N \cdot K$ valores de conexiones entre osciladores. En este trabajo los valores de d_i se determinan de antemano, por lo que quedan $N \cdot (2 + K)$ variables a ajustar.

Para encontrar valores adecuados para las variables se necesita realizar un proceso de optimización. En este trabajo se utiliza Optimización por Enjambre de Partículas (PSO, por sus siglas en inglés), método que usa una población para explorar el espacio de variables en búsqueda de una solución cercana al óptimo. Dado un tamaño de población, ésta se inicializa de manera aleatoria, aunque en rangos adecuados para que ocurran las dinámicas oscilatorias en los conjuntos KII. Se inicia el proceso de PSO con una cantidad de iteraciones máximas determinada. El rango de acción de las partículas es limitado para mantener las poblaciones válidas en todo momento. La evolución se realiza utilizando funciones de *fitness* sobre información contenida en la serie de tiempo original y en la serie de tiempo simulada mediante los modelos de Neuropercolación.

Un modelo de Neuropercolación permite realizar un análisis tanto en el dominio del tiempo como en el de frecuencia. En el tiempo, un modelo de Neuropercolación bien ajustado es útil para realizar interpolación y predicción, aunque tales características no son profundizadas en este trabajo. En frecuencia, la descomposición en señales simples facilita la identificación de frecuencias que, aunque presentes, puedan tener baja energía en el espectro de la señal original. En el caso de estudio de estrellas Gamma Doradus la identificación acertada de tales frecuencias puede llegar a ser de suma importancia para la determinación de parámetros físicos de la estrella. En la figura 3.1 se presenta un diagrama de la metodología ya explicada para ajustar los modelos de Neuropercolación a los datos.

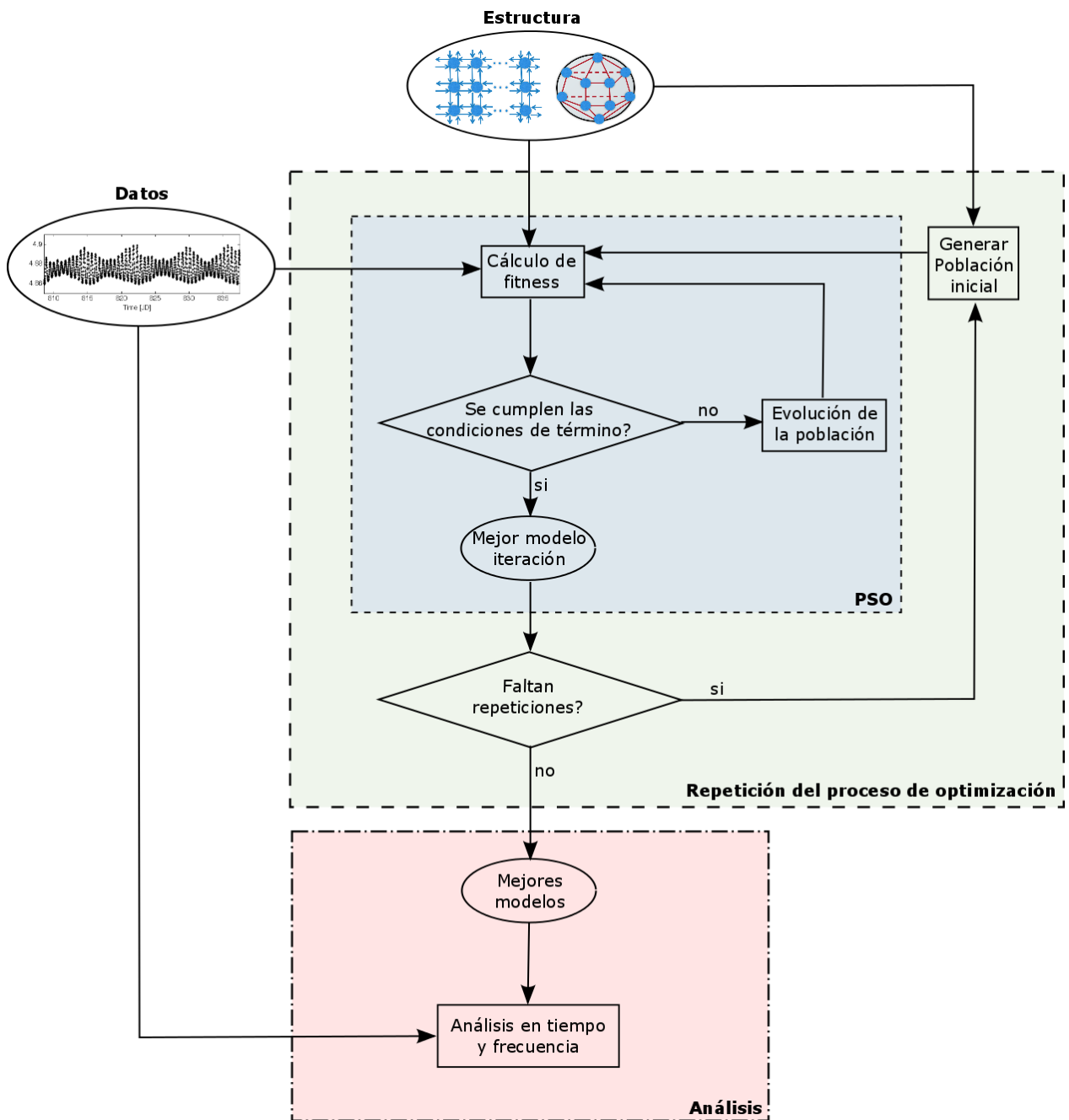


Figura 3.1: Metodología utilizada para encontrar los parámetros de un modelo de Neuro-percolación para un conjunto de datos. Primero se debe definir la estructura a utilizar. El proceso de optimización (PSO) es realizado múltiples veces para compensar la aleatoriedad del proceso.

3.1. Modelos

3.1.1. Modelos planos

Este tipo de modelos consisten en una grilla de osciladores ordenados de forma rectangular. Los osciladores son conectados entre ellos según un criterio de cercanía topológica, utilizando condiciones de borde toroidales. Este tipo de estructuras son de la forma mostrada en la figura 3.2(a), aunque la cantidad de osciladores puede variar. La cantidad de conexiones topológicas asociadas a cada oscilador es un parámetro a determinar, que en el caso particular de las pruebas realizadas en el presente trabajo fue fijado a cuatro vecinos como máximo. En el caso de estructuras de una sola fila, es posible definir únicamente dos vecinos (arriba y abajo). Para estructuras con dos filas y dos columnas (cuadradas) los vecinos corresponden a los osciladores de izquierda o derecha y arriba o abajo, sumando únicamente dos conexiones para cada oscilador. Cuando existen dos filas y tres o más columnas hay tres conexiones posibles: arriba o abajo, izquierda y derecha. Para estructuras de mayor tamaño, como la mostrada en la figura 3.2(a) se llega al máximo de cuatro conexiones por oscilador.

3.1.2. Modelos esféricos

De manera adicional a los modelos planos ya mencionados, se propone la utilización de modelos con una estructura similar a una esfera, tal como se observa en la figura 3.2(b). Para ello, se interconectan capas de cuatro osciladores, de manera similar a los modelos planos. Bajo esta configuración, los osciladores de los polos se encuentran conectados a todos los elementos de la capa que se encuentra inmediatamente bajo o sobre ellos. Las capas intermedias se conectan de manera similar a los modelos planos, con conexiones a los osciladores que se encuentran arriba, abajo, a la derecha e izquierda. Esta estructura se propone bajo el supuesto de que es similar a la que podría tener una estrella.

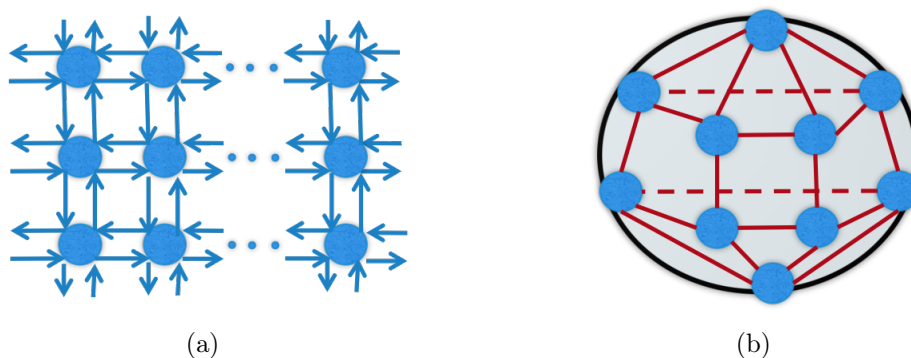


Figura 3.2: Estructuras utilizadas para los modelos de Neuropercolación. La figura 3.2(a) corresponde a los modelos planos con conexiones toroidales, donde cada oscilador tiene cuatro vecinos. La figura 3.2(b) corresponde a los modelos esféricos propuestos, que en este caso también presentan cuatro vecinos por oscilador.

3.2. Métodos de estimación de parámetros

Los parámetros ajustables del algoritmo son todas las probabilidades asociadas a cada oscilador: probabilidad de inhibición (p_i), probabilidad de seguir a la mayoría (ω_i) y las conexiones entre osciladores (c_{ji}). En general, las probabilidades de inhibición (p_i^-) y excitación (p_i^+) pueden ser distintas, pero por simplicidad en este trabajo se asume que $p_i^- = p_i^+ = p_i$. El tamaño de la vecindad (d_i) se fija *a priori*, y con el mismo valor para todos los osciladores, que en este caso ha sido fijado en $d = 11$. La elección de este valor se basa en criterios de observación de las dinámicas de los modelos. Para valores bajos de d la cantidad de frecuencias que pueden ser representadas por un conjunto KII es más pequeña y puede no ser suficiente para ciertos problemas. Más aún, se observa que para valores $d > 10$ no existe una mejora significativa en las frecuencias que se pueden generar mediante el modelo. Considerando que el aumento del valor de d conlleva un alza significativa en la carga computacional se definió el valor previamente mencionado $d = 11$.

3.2.1. Optimización por enjambre de partículas

Se decidió utilizar el método PSO para encontrar los parámetros adecuados para un modelo de Neuropercolación. Esto debido a que tales parámetros no se conocen *a priori*, sumado a que las ecuaciones de estos modelos son discretas recursivas y por ende se dificulta la utilización métodos de gradiente para la estimación de parámetros. Otros métodos de optimización pueden ser utilizados en lugar de PSO, manteniendo el resto de la metodología intacta.

En PSO hay diversos parámetros que deben ser ajustados. En cuanto a los parámetros asociados a la evolución de las partículas, éstos fueron fijados de antemano y no sufrieron cambios a lo largo de las pruebas realizadas. Se utilizó $\varphi_1 = \varphi_2 = 2$ e inercia que disminuye linealmente desde 0,95 hasta 0,4 durante el 70 % de las iteraciones. El tamaño de la población y las iteraciones máximas fueron escogidas dependiendo de las pruebas específicas realizadas.

El algoritmo puede tener diversas condiciones de término, siendo las más utilizadas las siguientes:

- **Por número de iteraciones:** Al llegar a un número máximo de iteraciones el proceso de optimización se detiene aunque no se haya encontrado el óptimo global. Esto permite que el proceso de optimización no continúe por un largo período de tiempo, a pesar de no haber convergido a la solución óptima.
- **Por convergencia:** Si el óptimo encontrado no cambia por un número determinado de iteraciones, entonces se dice que el proceso de optimización ha convergido a una solución, sea ésta local o global.

En este trabajo se utilizó únicamente el criterio de detención por iteraciones.

3.2.2. Manejo de restricciones

Debido a que se utiliza un modelo probabilístico, inherentemente hay restricciones relacionadas con los valores máximos y mínimos que puede tomar cada variable. En particular, se tiene que en cada oscilador G_i los parámetros que definen su dinámica y las conexiones con osciladores G_j presentan las siguientes restricciones:

$$p_i \in (0, 1) \quad \forall i \quad (3.1)$$

$$\omega_i \in (0, 1) \quad \forall i \quad (3.2)$$

$$c_{j,i} \in (0, 1) \quad \forall (i, j) \quad (3.3)$$

$$d \in \mathbb{N} \quad (3.4)$$

$$p_i + \sum_j c_{j,i} \leq 1 \quad (3.5)$$

En la implementación utilizada se considera relevante que todas las soluciones exploradas sean válidas. Esto significa que, al utilizar PSO, los individuos del enjambre deben tener limitada su área de movimiento al espacio incluido en las restricciones previamente mencionadas. En la práctica, estos límites se aplican en el momento de actualizar las posiciones de los individuos del enjambre. Para ello se ordenan las variables de forma aleatoria, de modo que no haya sesgo en el orden de éstas, y luego se actualiza la misma variable para toda la población. El rango de movimiento de cada partícula para tal variable se va a encontrar limitado por las restricciones descritas previamente, de la manera explicada en el pseudo-código del algoritmo 1.

Como se puede observar en el algoritmo dado y en las ecuaciones anteriores, algunas variables dependen de otras para el establecimiento de sus límites (en particular las conexiones entre osciladores y las probabilidades de inhibición, debido a la restricción de la ecuación 3.5). Por esta razón esa parte del establecimiento de los límites debe ser realizada de forma secuencial. Por otro lado, no es necesario que cada individuo de la población sea actualizado de forma secuencial, sino que se puede hacer paralelamente. En la implementación de esta metodología, para cada variable se actualizan simultáneamente todos los individuos de la población.

Algorithm 1 Algoritmo de manejo de restricciones en PSO

```
1: input Población de PSO:  $Xold$ 
2:  $nVar \leftarrow$  cantidad de variables a optimizar
3:  $nPop \leftarrow$  tamaño de la población
4:  $index \leftarrow$  permutación aleatoria de  $nVar$ 
5:
6: for  $j$  desde 1 a  $nPop$  do
7:   for  $i$  en  $index$  do
8:      $Xold(j,i) \leftarrow$  estado de la partícula  $j$  para la variable  $i$ 
9:      $vStep(j,i) \leftarrow$  velocidad de la partícula  $j$  para la variable  $i$  según PSO
10:     $Xnew(j,i) \leftarrow Xold(j,i) + vStep(j,i)$ 
11:    if variable  $i$  es probabilidad de inhibición then
12:       $c \leftarrow$  conexiones asociadas al oscilador
13:       $Xmin \leftarrow 0$ 
14:       $Xmax \leftarrow \min\{1, 1 - \text{suma}(c)\}$ 
15:    else if variable  $i$  es probabilidad de seguir a la mayoría then
16:       $Xmin \leftarrow 0$ 
17:       $Xmax \leftarrow 1$ 
18:    else if variable  $i$  es conexión entre osciladores then
19:       $cSum \leftarrow$  suma de conexiones asociadas al oscilador distintas de  $i$ 
20:       $p \leftarrow$  probabilidad de inhibición asociada al oscilador
21:       $Xmin \leftarrow 0$ 
22:       $Xmax \leftarrow \min\{1, 1 - p - cSum\}$ 
23:
24:    if  $Xnew(j,i) < Xmin$  then
25:       $Xnew(j,i) \leftarrow Xmin$ 
26:    else if  $Xnew(j,i) > Xmax$  then
27:       $Xnew(j,i) \leftarrow Xmax$ 
28:
29: return Nueva población de PSO:  $Xnew$ 
```

3.2.3. Creación de poblaciones iniciales

Para crear poblaciones iniciales válidas se aplican condiciones más restrictivas que las mencionadas en el manejo de restricciones, de manera que los individuos no se inicialicen en los límites de la región de parámetros válida, sino en la región central. De esta manera se utiliza que, para osciladores G_i y G_j los parámetros se inicializan en los siguientes rangos:

$$p_i \in (0,1; 0,5) \quad \forall i \quad (3.6)$$

$$\omega_i \in (0,5; 0,9) \quad \forall i \quad (3.7)$$

$$c_{i,j} \in (0; 0,05/K) \quad \forall (i, j). \quad (3.8)$$

3.2.4. Funciones de fitness

En el problema planteado se deben comparar dos series de tiempo uniformemente muestreadas, que pueden tener datos perdidos. De tal manera, la función de *fitness* utilizada en PSO para representar el ajuste de un individuo de la población, debe poder realizar esta comparación. Al trabajar con series de tiempo, dos caminos son posibles: comparar en el dominio del tiempo o en el dominio de la frecuencia.

Fitness en el dominio del tiempo

El beneficio de analizar las series en el dominio del tiempo es que no se requieren mayores transformaciones, y por lo tanto no aumenta la complejidad computacional. El primer ejemplo de método de comparación en el dominio del tiempo sería el Error Cuadrático Medio (MSE por sus siglas en inglés) que se define como

$$MSE = \frac{1}{N} \sum_{i=1}^N (X_i - \hat{X}_i), \quad (3.9)$$

donde X_i corresponde a la serie de tiempo original y \hat{X}_i corresponde a su estimación. La cantidad de puntos de la serie de tiempo a comparar es N . Estos puntos deben pertenecer al mismo instante de tiempo en la serie. Se pueden realizar mejoras de este método mediante ponderaciones de cada punto, por ejemplo, utilizando el error de medición. De esta manera se penaliza en menor grado si la estimación se encuentra alejada del valor medido cuando tal medición no es confiable.

Para poder realizar esta comparación de forma efectiva es necesario que ambas series de tiempo a comparar se encuentren en la misma escala. Esto es relevante porque la salida de un modelo de Neuropercolación son probabilidades de activación, las cuales se encuentran entre 0 y 1, mientras que la serie de tiempo que se quiere modelar es de carácter arbitrario. Debido a esto, previo al cálculo del Error Cuadrático Medio se normalizan ambas series de tiempo para que tengan valor medio 0 y varianza 1.

En el caso del problema abordado en este trabajo, la señal original y el modelo no comparten un punto de inicio. Por ello, a la hora de calcular el MSE, la señal modelada se desplaza en un rango de instantes de tiempo, de manera de encontrar el mejor ajuste entre estos desplazamientos.

Fitness en el dominio de la frecuencia

Otra opción es el uso de comparaciones de espectros de frecuencia. Esto añade complejidad al problema y debe ser tratado cuidadosamente, dado que se debe escoger una representación adecuada en frecuencia y además una buena forma de comparar las representaciones de dos señales.

3.3. Datos a utilizar

Los datos a utilizar para probar la metodología propuesta se dividen en dos conjuntos: datos sintéticos y datos reales. Los primeros consisten de un conjunto de series de tiempo simuladas que representan diversos fenómenos de interés. Entre estos se encuentran modulaciones de amplitud, señales periódicas no sinusoidales, sumas de señales periódicas con períodos inconmensurables, entre otros. Para el caso de las señales reales se utilizaron datos del proyecto Kepler de la NASA [21], el cual mide la luminosidad de estrellas en la Vía Láctea con el objetivo de encontrar sistemas planetarios.

3.3.1. Datos sintéticos

Al utilizar datos sintéticos es importante considerar que éstos tienen que contener las características que se desean estudiar, que en el presente trabajo corresponde a la cuasi-periodicidad. Esta dinámica puede ser reconstruida de diversas maneras, entre ellas las que fueron utilizadas en este trabajo:

- a. **Modulación de amplitud:** Dependiendo de las señales, y en particular las frecuencias, que se utilicen al definir una modulación de amplitud, ésta puede llevar a características cuasi-periódicas. Esto se debe a la equivalencia que existe entre la multiplicación de dos señales sinusoidales con sumas de sinusoides.
- b. **Modelos de estrellas Gamma Doradus:** Debido a que se sabe que las estrellas Gamma Doradus presentan un comportamiento cuasi-periódico, y el objetivo final es analizar las propiedades de la metodología planteada en señales reales, estudiar modelos de este tipo de estrellas corresponde a un paso importante para establecer los alcances de esta propuesta. Para definir un modelo de estrella en el tiempo se obtienen frecuencias (o períodos) factibles mediante modelos físicos, y luego se aplican estas frecuencias en modelos analíticos empíricos que emulan la variación de luminosidad de este tipo de estrellas. Van Reeth et al. [20] presenta un modelo empírico de curva de luz de Gamma Doradus, acompañado por conjuntos de frecuencias admisibles para tal tipo de estrella.

3.3.2. Datos del proyecto Kepler

El proyecto Kepler fue una misión que tuvo como objetivo explorar la estructura y diversidad de sistemas planetarios extrasolares [21]. Este proyecto de la NASA (National Aeronautics and Space Administration) consiste en un satélite con un telescopio en su estructura. Este telescopio comenzó su funcionamiento en Mayo de 2009, durando hasta Mayo de 2013. Su búsqueda se realizó en regiones de las constelaciones de Cygnus y Lyra en nuestra galaxia, la Vía Láctea (ver figura 3.3.2). Hasta la fecha la misión ha confirmado 2326 planetas extrasolares, teniendo aún 4706 candidatos.

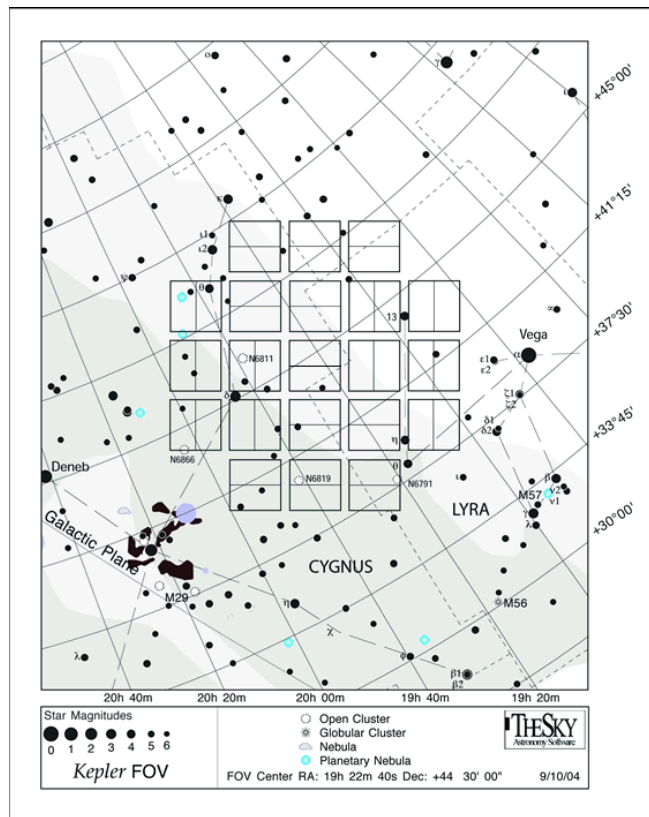
La misión Kepler tuvo que terminar debido a fallas técnicas del satélite: dos de sus ruedas de reacción, que permitían ajustar la dirección de observación, dejaron de funcionar. Pese a ello, la misión K2 fue anunciada y comenzó su funcionamiento en Marzo del 2014 [57]. Esta segunda misión utiliza el mismo satélite y telescopio, con un nuevo enfoque que permite utilizar las potencialidades del satélite, pese a sus fallas, apuntando hacia los objetos que se observan secuencialmente mientras realiza su órbita en torno al sol.

La misión Kepler mantiene sus datos públicos, siendo liberados cuatro meses después del término de cada ciclo de observación. Los datos son liberados en distintos formatos, incluyendo imágenes completas de los espacios del cielo observado y curvas de luz de las estrellas que se observan. Los datos fotométricos obtenidos por el proyecto Kepler son de dos naturalezas distintas, en cuanto a la regularidad con la cual se adquieren los datos: la *Cadencia Larga* (LLC) y la *Cadencia Corta* (SLC). La primera tiene un periodo entre datos de aproximadamente media hora (0.49 h), mientras que la segunda presenta datos en cada minuto (58.85 s). En este trabajo, debido a las escalas de tiempo con las que se trabajan, los datos de cadencia larga son suficientes para el análisis por lo que no se utilizan los datos de cadencia corta.

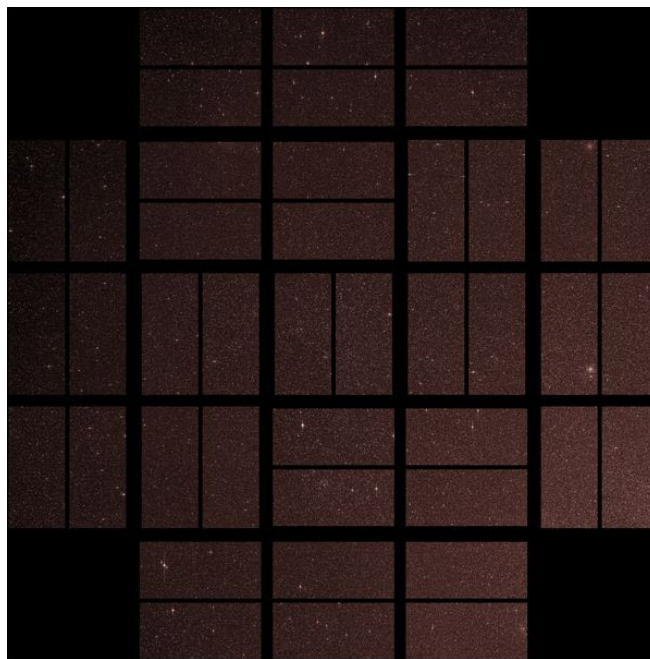
Los datos públicos de Kepler se dividen en *quarters*, que son períodos distintos de la misión (ver tabla 3.1). Datos tales como la calidad general de los datos obtenidos en cada *quarter*, además de eventos anómalos sucedidos durante las mediciones pueden encontrarse en el documento *Kepler Data Release* [58]. En particular, los *quarters* 4 y 12 presentan anomalías que hacen que sea deseable evitar su uso en este trabajo.

Quarter	Inicio	Término	Quarter	Inicio	Término
0	2009 May 02	2009 May 11	9	2011 Mar 19	2011 Jun 27
1	2009 May 13	2009 Jun 15	10	2011 Jun 28	2011 Sep 27
2	2009 Jun 20	2009 Sep 16	11	2011 Sep 29	2012 Jan 04
3	2009 Sep 18	2009 Dec 16	12	2012 Jan 05	2012 Mar 28
4	2009 Dec 19	2010 Mar 19	13	2012 Mar 29	2012 Jun 27
5	2010 Mar 20	2010 Jun 23	14	2012 Jun 28	2012 Oct 03
6	2010 Jun 24	2010 Sep 22	15	2012 Oct 05	2013 Jan 11
7	2010 Sep 23	2010 Dec 22	16	2013 Jan 12	2013 Apr 08
8	2010 Dec 22	2011 Mar 24	17	2013 Apr 09	2013 May 08

Tabla 3.1: *Quarters* de datos del proyecto Kepler y las fechas entre las cuales fueron tomados los datos.



(a)



(b)

Figura 3.3: Área de observación del telescopio espacial Kepler. Incluye secciones de las constelaciones Cygnus y Lyra en la Vía Láctea. La figura 3.3(a) corresponde a un esquema de su ubicación, y 3.3(b) corresponde a una observación real del telescopio [56].

En este trabajo se utilizan datos fotométricos del catálogo KIC (*Kepler Input Catalog*) [59] para aquellas estrellas clasificadas como Gamma Doradus en el CDS (*Strasbourg astronomical Data Center*) [22]. Esto permite obtener un catálogo de 199 estrellas con sus respectivas curvas de luz, de las cuales se analizan 3 elegidas por ser representativas de las dinámicas que motivan este trabajo, así como por presentar mayor cantidad de información adicional. Las curvas de luz se obtienen del archivo MAST (Mikulski Archive for Space Telescopes) [60]. Cada curva de luz de una estrella, para un quarter definido se presenta en un archivo formato FITS (Flexible Image Transport System), que es estándar para datos astronómicos. En este archivo se encuentra la imagen de la estrella, la información sobre la fotometría y metadata. La información fotométrica incluyen datos crudos, así como datos ya procesados por la *pipeline* de Kepler.

El tiempo entregado por Kepler corresponde al *Baricentric Kepler Julian Date* (Día Juliano baricéntrico de Kepler), o BKJD. Esta medida corresponde a la cantidad de días transcurridos desde el mediodía del 1 de Enero de 2009, corregidos según el movimiento de la nave alrededor del centro de masa del sistema solar. Una estimación adecuada del tiempo es necesaria para obtener información de períodos estelares de la manera más precisa posible. Según el Manual de Archivo de Kepler [61], se cree que los tiempos para cualquier cadencia tienen una precisión de $\pm 50ms$. Adicionalmente es sabido que el procesamiento utilizado por Kepler produce artefactos en los datos [62], por lo que algunos estudios realizan su propio procesamiento de los datos (eliminación de tendencias y fallas de medición).

La información sobre la fotometría incluye tanto el flujo sin correcciones (SAP_FLUX) como el flujo corregido de acuerdo a la *pipeline* de Kepler (PDCSAP_FLUX). Cada uno de estos flujos es acompañado por el error fotométrico estimado. El flujo sin correcciones corresponde al flujo en unidades de electrones por segundo en la apertura óptima. Su error incluye tanto errores instrumentales como de cálculo. El flujo corregido incluye el procesamiento para eliminar las tendencias de las curvas de luz asociadas a las mediciones.

Con la información previamente mencionada se construye una curva de luz utilizando el BKJD, el flujo corregido (PDCSAP_FLUX) y su respectivo error (PDCSAP_FLUX_ERR). A los datos fotométricos se les aplica una transformación logarítmica en base 10. Este proceso se realiza debido a que en astronomía la luminosidad de una estrella es medida en magnitud aparente con respecto a una referencia, la cuál está dada por la siguiente transformación,

$$m_{obj} - m_{ref} = -2,5 \log_{10} \left(\frac{I_{obj}}{I_{ref}} \right), \quad (3.10)$$

donde I_{obj} y I_{ref} son las luminosidades medidas para un objeto específico y para uno de referencia para el cual la magnitud ya es conocida (m_{ref}). Debido a que no se posee información sobre una estrella u objeto brillante que pueda usarse como referencia, la transformación logarítmica es suficiente para obtener la forma de la curva de luz del objeto de interés. Algunos ejemplos de curvas de luz de estrellas Gamma Doradus, después de realizado este proceso, se ilustran en la figura 3.4.

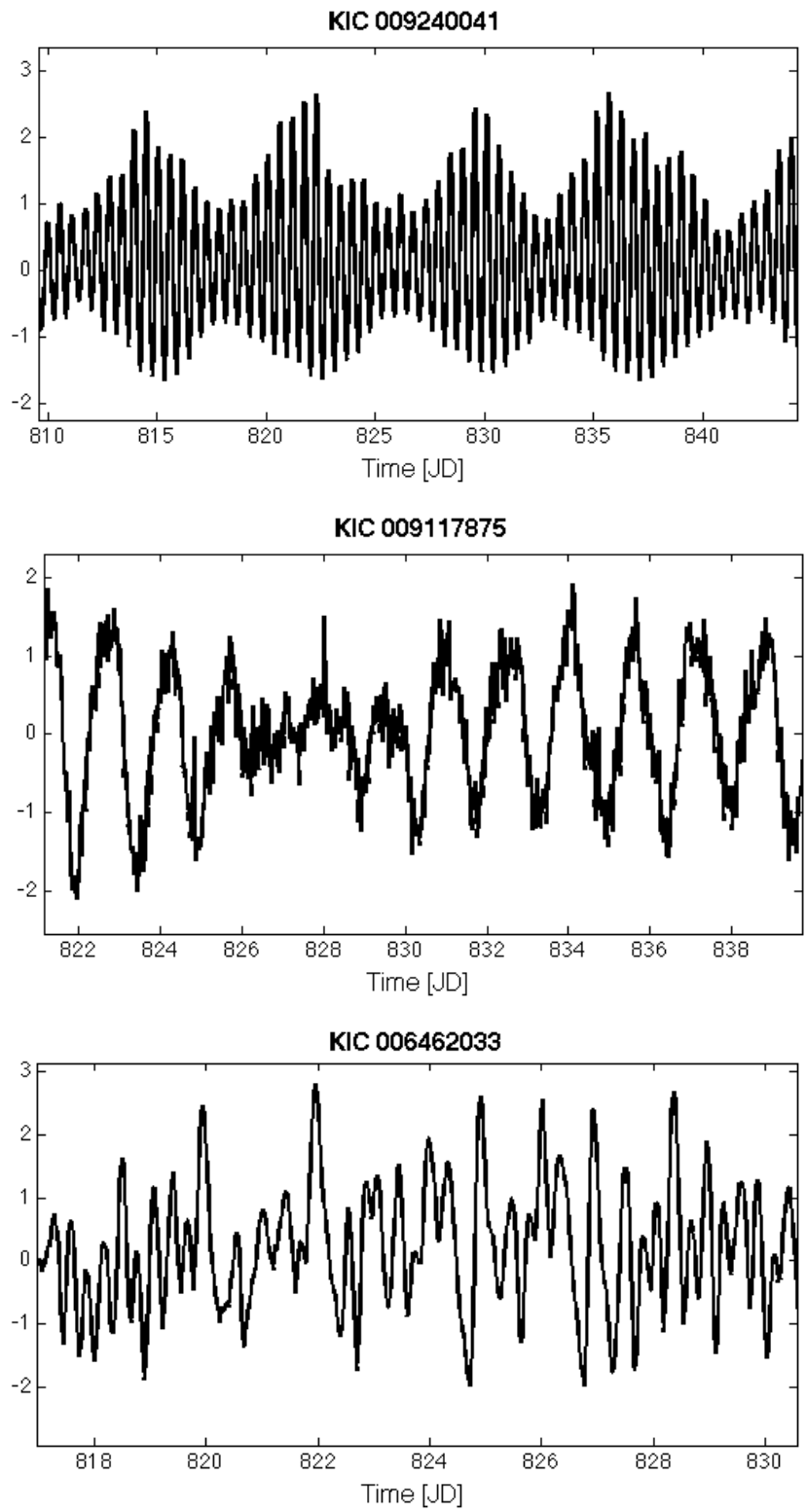


Figura 3.4: Ejemplos de curvas de luz tomadas de Kepler de estrellas Gamma Doradus después del pre-procesamiento

3.4. Herramientas de análisis de resultados

El resultado que entrega la metodología propuesta corresponde a una serie de tiempo uniformemente muestreada en el rango de tiempo de los datos que se desea modelar. Se tienen tres objetivos principales a la hora de analizar las propiedades de los modelos obtenidos:

- **Analizar la bondad de ajuste** determinada por la función escogida para el *fitness* de la población en PSO. Esta función tiene que ser capaz de comparar dos series de tiempo evaluadas en los mismos instantes de tiempo, y que puede tener datos perdidos. En el caso del trabajo aquí presentado se escogió Error Cuadrático Medio como la medida de ajuste, aunque otras medidas en tiempo o frecuencia pueden ser utilizadas. Se escogió el MSE debido a que se trabajó con datos con poco o nulo nivel de ruido gaussiano.
- **Probar la robustez del estimador.** Esto debe realizarse debido a que los resultados de PSO tienen una variabilidad intrínseca debido a la aleatoriedad con la que se crean las poblaciones iniciales y con que los individuos de la población se mueven en el espacio de características. Por esta razón, una sola iteración de PSO puede entregar valores que no sean representativos de su comportamiento promedio. Se analizan los resultados de mínimo error, el error promedio y la mediana, para poder tener una visión general de la distribución de los resultados.
- **Evaluar los beneficios del modelo relacionados a la identificación de períodos de oscilación,** lo cual es particularmente relevante en el caso de estudio presentado. No basta encontrar un modelo que se ajuste perfectamente a los datos si no entrega una mejora relativa a su interpretabilidad. Para este estudio se analizarán los períodos principales asociados a cada oscilador de un modelo dado, de manera de observar si permite la identificación de períodos relevantes que pasan desapercibidos entre el ruido del periodograma.

Con estos objetivos en mente, para poder realizar el análisis de los resultados obtenidos, se realizan múltiples realizaciones independientes de PSO para cada caso a analizar, reportando sus ajustes (MSE) mínimo, medio y mediano. Luego, los mejores resultados se descomponen en series de tiempo correspondientes a los osciladores. Las frecuencias más importantes de cada oscilador, según la energía que presentan en el espectro, entregan una idea de cómo se divide cada componente de frecuencia en la estructura. Todos los análisis en frecuencia se realizan utilizando periodograma de Lomb-Scargle, debido a que admite datos faltantes y además puede ajustarse el factor de sobre-completitud del diccionario.

3.5. Programación

La implementación principal de este método fue realizada en el programa MATLAB, utilizando el Toolbox de *Parallel Computing* para acelerar los cálculos. Las implementaciones de los códigos más importantes en lenguaje MATLAB se encuentran en el anexo 7.2.

4 Resultados

4.1. Pruebas realizadas

Para validar los supuestos realizados en este trabajo se utiliza la metodología ya presentada sobre diversos conjuntos de datos, de manera de abarcar múltiples dinámicas cuasi-periódicas. Estos conjuntos se encuentran compuestos de datos sintéticos y reales, lo que permite el estudio de la metodología desde distintos ángulos. Los datos sintéticos permiten estudiar las propiedades del método y los modelos, debido a que se conocen las características en tiempo y frecuencia que se desean obtener; mientras que los datos reales presentan un desafío mayor al no conocerse sus características con certeza, pero asimismo ofrecen opciones de interpretabilidad real de los resultados.

El primer conjunto de datos sintéticos utilizado corresponde a un modelo simple de una señal de Amplitud Modulada. Con conocimiento de que es posible modelar esta señal de manera precisa utilizando Neuropercolación, se plantean pruebas diseñadas para estudiar la robustez de la estimación. En primer lugar se analiza el efecto que tiene la pérdida de datos en los modelos obtenidos, para luego estudiar la influencia de la presencia de ruido blanco aditivo en éstos.

El segundo conjunto de datos sintéticos corresponde a curvas de luz simuladas de estrellas Gamma Doradus. Estos modelos son obtenidos mediante la utilización de las frecuencias entregadas por Van Reeth et al. [20]. En este caso se estudian diferentes configuraciones de modelos de Neuropercolación, con el objetivo de definir el mejor modelo. Se estudia la distribución de los períodos encontrada, comparándola con el objetivo deseado y discutiendo los beneficios y dificultades que este esquema presenta a la hora de analizar el espaciamiento de períodos.

Finalmente, se estudian curvas de luz reales con datos provenientes del proyecto Kepler. Las estrellas observadas representadas por estas curvas de luz han sido confirmadas como Gamma Doradus por diversas fuentes, apareciendo en CDS con tal etiqueta. Los identificadores de Kepler para las estrellas estudiadas en este punto son KIC 9240041, KIC 10224094 y KIC 11145123. Para el análisis de estas curvas de luz se ajustan diferentes modelos de Neuropercolación, escogiendo y estudiando en profundidad los que obtienen un mejor ajuste. Se pone especial énfasis a los períodos encontrados y a los residuos obtenidos, que ilustran los problemas que presentan los modelos.

4.2. Señal AM Sintética

La portadora es generada con una frecuencia de $f_c = \sqrt{2}$, mientras que la señal de información presenta una frecuencia de $f_i = 0,05$,

$$x(t) = \sin\left(2\pi\sqrt{2}t\right) \cos\left(2\pi 0,05t\right). \quad (4.1)$$

Se generó una serie de tiempo de 2000 puntos utilizando la ecuación 4.1 con una frecuencia de muestreo de 1/48 días, emulando el muestreo realizado por la misión Kepler en su cadencia larga. La señal y su diagrama de fase se pueden observar en la figura 4.1. Al analizar esta señal se espera que se puedan encontrar osciladores asociados a la suma y resta de las frecuencias debido a que esta señal de amplitud modulada tiene como expresión equivalente la de la ecuación 4.2 ,

$$x(t) = \frac{1}{2} \left[\cos\left(2\pi(\sqrt{2} + 0,05)t - \frac{\pi}{2}\right) + \cos\left(2\pi(\sqrt{2} - 0,05)t - \frac{\pi}{2}\right) \right]. \quad (4.2)$$

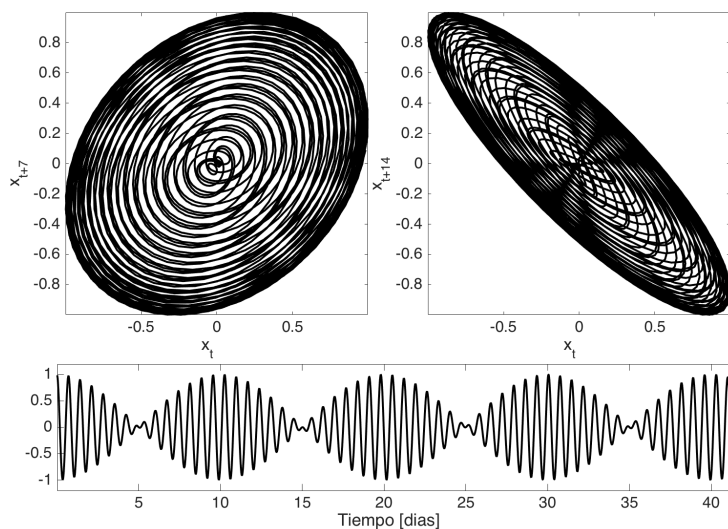


Figura 4.1: Ejemplo de una señal de amplitud modulada simulada, ilustrada en el gráfico inferior. El gráfico de fase a la izquierda muestra cómo luce el atractor cuando la dimensión del embedding es $M = 2$ y el delay es $\tau = 7$, mientras que el de la derecha usa $\tau = 14$.

Por medio de conocimiento adquirido en pruebas preliminares, se escoge una estructura con cuatro osciladores ordenados en una grilla de dos por dos para la modelación. En esta configuración cada oscilador presenta dos conexiones: hacia el lado y hacia arriba/abajo. Se observó que estructuras tan pequeñas como dos osciladores son capaces de representar este tipo de señales, pero el grafo con cuatro elementos entrega una mayor flexibilidad para la optimización, alcanzando por lo tanto resultados más confiables. Al existir más osciladores que frecuencias a representar, es posible encontrar redundancias o osciladores que no muestren movimiento. Sin embargo no se desea encontrar frecuencias principales distintas a las que aparecen en la ecuación 4.2.

4.2.1. Robustez ante datos perdidos

Para conocer el efecto que producen los datos perdidos en la estimación del modelo, y asimismo descubrir la proporción de puntos faltantes que puede manejar el método, se reduce la cantidad de datos presentada y se analizan los modelos obtenidos. Partiendo con el conjunto completo de datos (100 %) como referencia, se reduce la cantidad de información en la serie de tiempo en intervalos de 10 % hasta utilizar únicamente el 10 % de los datos. Los puntos se extraen de la serie de tiempo mediante un algoritmo de selección aleatoria independiente. Cada experimento fue repetido veinte veces utilizando inicializaciones aleatorias diferentes para las poblaciones de PSO.

Al momento de ejecutar la optimización mediante PSO, el valor de la función de *fitness* se calcula utilizando la serie de tiempo con datos faltantes. Posteriormente, después de obtener el mejor resultado de la optimización, también se calcula el ajuste (en este caso, MSE) para la serie de tiempo original. Este cómputo se realiza debido a que permite una comparación justa de los modelos obtenidos con datos diferentes. Los resultados de ambos cálculos se encuentran en detalle en la tabla 4.1.

Se observa que en todos los conjuntos de simulaciones se obtienen tanto resultados con un ajuste pobre, como también algunos muy acertados. Es importante destacar que los malos resultados ocurren cuando sólo una de las frecuencias buscadas es representada en el modelo, por lo que es razonable asumir que se trata de un mínimo local en el espacio de variables a optimizar. Por otro lado, aún cuando se usa sólo el 10 % de los datos (200 puntos en el caso de estudio), es posible modelar la señal de manera bastante exacta, tal como se puede observar en la figura 4.2.

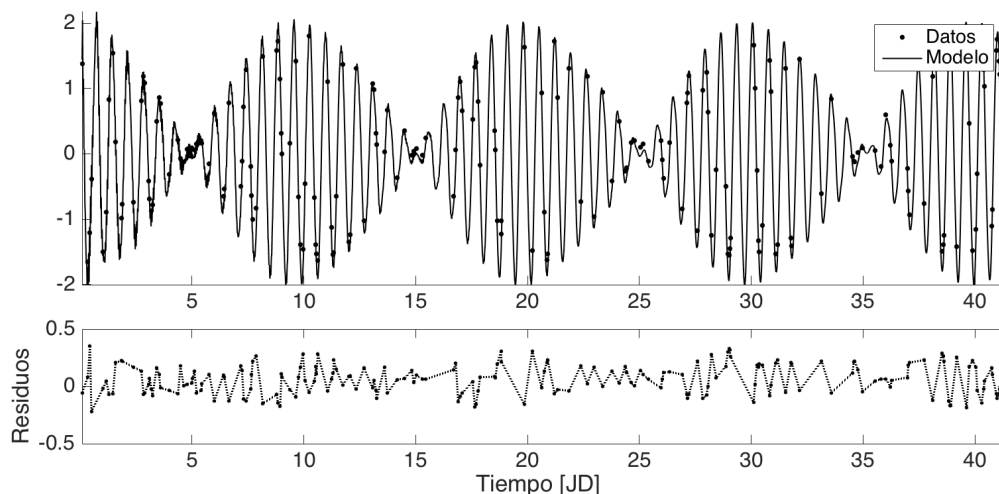


Figura 4.2: Modelo de Neuropercolación para una señal AM sintética con datos perdidos. El modelo fue obtenido usando un 10 % de los datos originales, y aún así captura la dinámica de la señal original. El gráfico inferior muestra los residuos entre la señal original y el modelo.

Tabla 4.1: MSE de los modelos de Neuropercolación para una señal AM sintética cuando se utiliza información incompleta para la estimación. Los valores de fitness son calculados utilizando el conjunto de datos completo y el de datos con información faltante.

Datos utilizados [%]	Señal con datos perdidos			Señal original		
	Mínimo	Mediana	Media	Mínimo	Mediana	Media
10	0.0206	0.1715	0.1939	0.0024	0.2231	0.2950
20	0.0016	0.4585	0.3008	0.0004	0.4485	0.3068
30	0.0034	0.2636	0.2673	0.0010	0.2815	0.2796
40	0.0034	0.4980	0.3058	0.0022	0.5466	0.3355
50	0.0032	0.2969	0.2805	0.0022	0.2953	0.2785
60	0.0011	0.2743	0.2710	0.0010	0.2840	0.2795
70	0.0006	0.1083	0.2528	0.0004	0.1151	0.2593
80	0.0021	0.0137	0.1978	0.0018	0.0132	0.1978
90	0.0005	0.5418	0.3540	0.0005	0.5448	0.3564
100	0.0011	0.0083	0.1845	0.0011	0.0083	0.1845

Tabla 4.2: Períodos principales de los mejores modelos de Neuropercolación para una señal AM sintética con información incompleta. Estos periodos se encuentran analizando cada oscilador por separado utilizando Lomb-Scargle. Se considera sólo el período principal de cada uno de los cuatro osciladores, siempre y cuando su potencia en el periodograma supere al umbral dado por una probabilidad de falsa alarma del 0.001 %. El único periodo distinto a lo esperado corresponde al oscilador 1 al utilizar el 10 % de los datos.

Datos utilizados [%]	Oscilador 1	Oscilador 2	Oscilador 3	Oscilador 4
10	0.0439	0.7319	0.6835	0.6835
20	0.6842	0.7327	0.6842	0.7327
30	0.6824	0.7335	0.6824	0.7335
40	—	0.7331	0.6820	0.6820
50	0.7338	0.7338	0.6827	0.6827
60	0.7342	0.7342	—	0.6831
70	0.7342	0.6831	0.7342	0.6831
80	0.6827	—	0.7338	0.7338
90	0.7342	0.6831	0.7342	0.6831
100	0.6831	0.6831	0.7342	—

Por otro lado, los períodos que representan los osciladores de los modelos encontrados se encuentran en el conjunto de valores esperados. Para definir los períodos de los osciladores se consideran sólo los máximos globales de los periodogramas, siempre y cuando aquellas frecuencias superen el umbral definido por una probabilidad de falsa alarma de 0.001 %. Se realiza esta distinción para evitar considerar frecuencias espurias encontradas en los conjuntos KII que no presentan oscilaciones. El detalle de los períodos encontrados en cada uno de los osciladores se encuentra en la tabla 4.2.

En la figura 4.3 se representan los periodos obtenidos mediante Lomb-Scargle de todos los osciladores en los mejores modelos encontrados. Es fácil ver que los valores de las frecuencias se encuentran cercanos únicamente a las dos componentes esperadas, excepto en el caso del mejor modelo utilizando el 10% de los datos. En tal caso, el Oscilador 1 presenta una componente de bajo periodo.

Incluso en los peores modelos obtenidos se encuentra al menos una de las frecuencias relevantes, lo que se puede apreciar en la figura 4.4. En todos los casos observados, al menos una de las frecuencias es representada en el modelo de Neuropercolación al finalizar el proceso de optimización. Esto podría deberse a la existencia de óptimos locales. En la figura señalada, también se observa que algunos modelos presentan componentes de alto periodo (cercano a 166 días) y bajo periodo (cercano a 0.05 días).

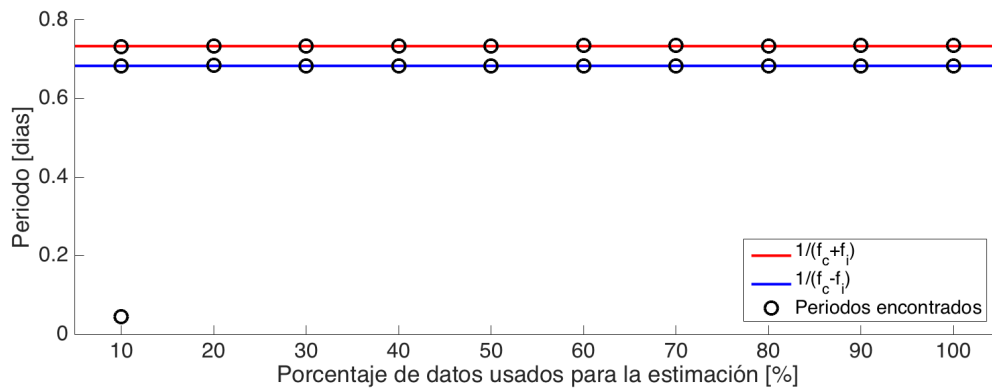


Figura 4.3: Períodos principales encontradas en los modelos de Neuropercolación obtenidos con señal AM sintética con información incompleta.

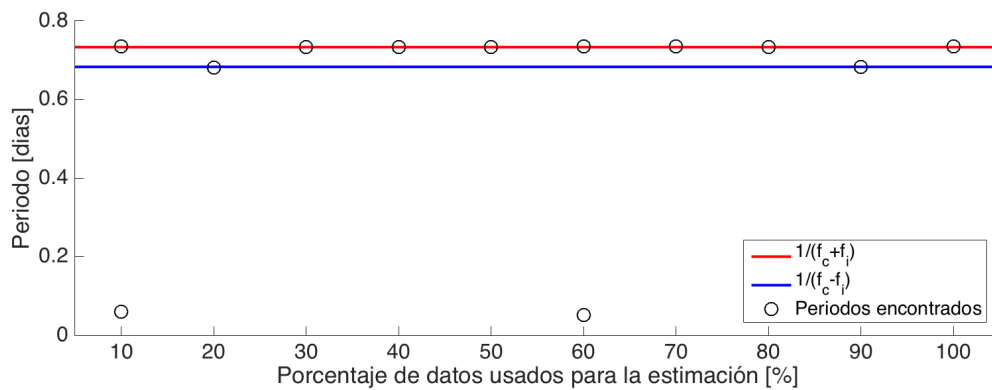


Figura 4.4: Períodos principales encontradas en los peores modelos obtenidos con señales AM sintética con información incompleta. Utilizando 10, 30 y 50% de los datos también se encontró una componente de largo período cercano a los 166 días.

4.2.2. Robustez al ruido

El método fue probado con seis niveles de ruido blanco gaussiano aditivo, correspondientes a valores de Señal a Ruido (SNR) de $\{30, 20, 15, 10, 5\}$, tal como se muestra en la ecuación 4.3,

$$x(t) = \sin\left(2\pi\sqrt{2}t\right) \cos(2\pi 0,05t) + \eta(t, \sigma), \quad (4.3)$$

donde $\eta(t, \sigma)$ representa el ruido blanco gaussiano aditivo de media cero y desviación estándar σ . Adicionalmente, la Señal a Ruido es computada como

$$SNR_{dB} = 10 \log_{10} \left(\frac{A_s}{A_n} \right)^2, \quad (4.4)$$

donde A_s y A_n son las amplitudes RMS (*Root Mean Square* en inglés) de la señal y del ruido, respectivamente.

Para cada nivel de ruido se utiliza la misma configuración del modelo, en este caso una grilla de dos por dos (cuatro osciladores). Nuevamente, la optimización es repetida veinte veces debido a la aleatoriedad intrínseca del método de optimización PSO. Al momento de realizar la optimización, el Error Cuadrático Medio es calculado para comparar la salida del modelo con la señal con ruido. Sin embargo, después de obtener los mejores modelos también se analiza cómo se comparan éstos con la señal original libre de ruido. Los resultados se encuentran detallados en la tabla 4.3. Como es de esperarse, el MSE al comparar los modelos con las señales ruidosas aumenta junto al nivel de ruido. Aún así, cuando los mejores modelos son comparados con la señal original, sin ruido aditivo, los valores de MSE disminuyen notoriamente.

En la figura 4.5 se tiene un diagrama de caja de los resultados. El aumento del MSE al comparar los modelos con las señales ruidosas es consistente con el nivel de ruido aditivo. Al comparar con la señal original, se obtienen rangos similares de MSE para los modelos obtenidos con diferentes niveles de ruido.

Tabla 4.3: MSE de los modelos de Neuropercolación para una señal AM sintética obtenidos cuando se le añade ruido blanco. Se incluyen los datos del MSE entre los modelos y las señales original y ruidosa.

SNR	Señal original			Señal con ruido		
	Mínimo	Mediana	Media	Mínimo	Mediana	Media
30	0.0002	0.0042	0.1561	0.0013	0.0052	0.1569
20	0.0007	0.0185	0.2236	0.0105	0.0281	0.2308
15	0.0011	0.0047	0.1422	0.0310	0.0338	0.1696
10	0.0045	0.5424	0.3752	0.0924	0.6001	0.4429
5	0.0031	0.5424	0.3330	0.2452	0.6913	0.5180
1	0.0015	0.0120	0.2506	0.4656	0.4710	0.6429

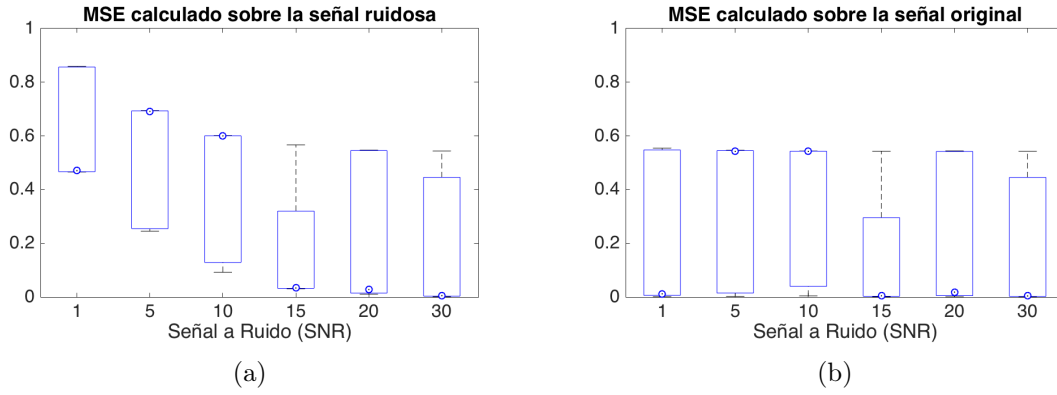


Figura 4.5: MSE de los modelos obtenidos al añadir ruido a una señal de amplitud modulada sintética. La figura 4.5(a) refleja el MSE al comparar los resultados con la señal ruidosa, que es la que se usa en el proceso de optimización. La figura 4.5(b) muestra el MSE al comparar con la señal original, sin ruido. Las cajas representan el rango entre el primer y tercer cuartil, mientras que los puntos señalan la mediana.

Tabla 4.4: Períodos principales de los modelos de Neuropercolación para una señal AM sintética con ruido blanco. Se obtienen utilizando Lomb-Scargle. Se considera sólo el máximo global en el periodograma de cada uno de los cuatro osciladores, siempre y cuando su potencia supere al umbral dado por una probabilidad de falsa alarma del 0.001%. Para los casos con mayor nivel de ruido (SNR 5 y 1) uno de los osciladores presenta una componente de periodo largo y otro contiene una componente de bajo periodo.

SNR	Oscilador 1	Oscilador 2	Oscilador 3	Oscilador 4
30	0.7342	0.6831	0.7342	0.6831
20	0.7342	0.6831	—	0.6831
15	0.7342	0.7342	0.6831	0.6831
10	—	—	0.6831	0.7310
5	55.5556	0.7310	0.0527	0.6831
1	0.7310	0.1068	0.6831	55.5556

Incluso cuando el nivel de ruido es alto, los períodos de los osciladores se encuentran bastante cercanas a los períodos buscados. Esto se puede observar en la tabla 4.4. Pese a ello, a mayor ruido aparecen componentes adicionales, tales como una de alta y otra de baja frecuencia en los casos de SNR 5 y 1.

4.3. Simulaciones de curvas de luz de estrellas Gamma Doradus

Utilizando los valores de frecuencias que Van Reeth et al. [20] reportan para diversos modos de oscilación de una estrella, se construyen curvas de luz sintéticas mediante la asignación de amplitudes y fases aleatorias a sumas de sinusoides con tales frecuencias. Las frecuencias reportadas por Van Reeth corresponden a modos de oscilación determinados por los números cuánticos l y m para un modelo de estrella Gamma Doradus. Las curvas de luz fueron creadas siguiendo la metodología del mismo trabajo:

$$F(t) = A \left(\left[1 + \sum_i a_i \sin(2\pi f_i t + \phi_i) \right]^{2,2} - 1 \right) + \sigma. \quad (4.5)$$

En la ecuación 4.5 A corresponde a la amplitud total de la curva simulada, a_i corresponde al factor de escala para el modo de oscilación $n = i$, ϕ_i es la fase de tal modo y σ representa el ruido aditivo presente. La potencia 2.2 se utiliza para otorgarle asimetría a la curva de luz, que es una característica muy presente en las variaciones de luminosidad de las estrellas Gamma Doradus. Las frecuencias f_i son las obtenidas de CDS. Para este trabajo se consideró $A = 1$ y ruido aditivo nulo. Las amplitudes relativas y fases fueron determinadas de manera aleatoria.

Para probar la metodología propuesta se considera un conjunto conformado por ocho frecuencias correspondientes al modo $l = 2$, $m = 2$ de una estrella Gamma Doradus simulada. Para la simulación de la serie de tiempo se emula el muestreo utilizado en Kepler, con intervalos de media hora entre muestras, generando una serie de tiempo de 2000 puntos para realizar la modelación. En este caso las frecuencias utilizadas, así como sus amplitudes y fases se muestran en la tabla 4.5. En la figura 4.6 se observa el espacio de fase que representa un atractor de tres dimensiones. Tal atractor es muy similar a algunos de los que se observan en estrellas Gamma Doradus reales, por ejemplo KIC 9240041, que es una de las que se estudia en este trabajo.

Tabla 4.5: Parámetros utilizados para la simulación de una curva de luz de estrella Gamma Doradus. Las frecuencias fueron obtenidas de los reportes de Van Reeth et al. (2015) en CDS. Las amplitudes y fases fueron determinadas de manera aleatoria.

Frecuencia	Amplitud	Fase
1.2444	0.1888	6.0162
1.2762	0.2099	6.0626
1.3031	0.0294	0.9903
1.3208	0.2116	6.0984
1.3386	0.1465	6.0141
1.3662	0.0226	3.0497
1.3997	0.0645	5.0283
1.4339	0.1267	0.8915

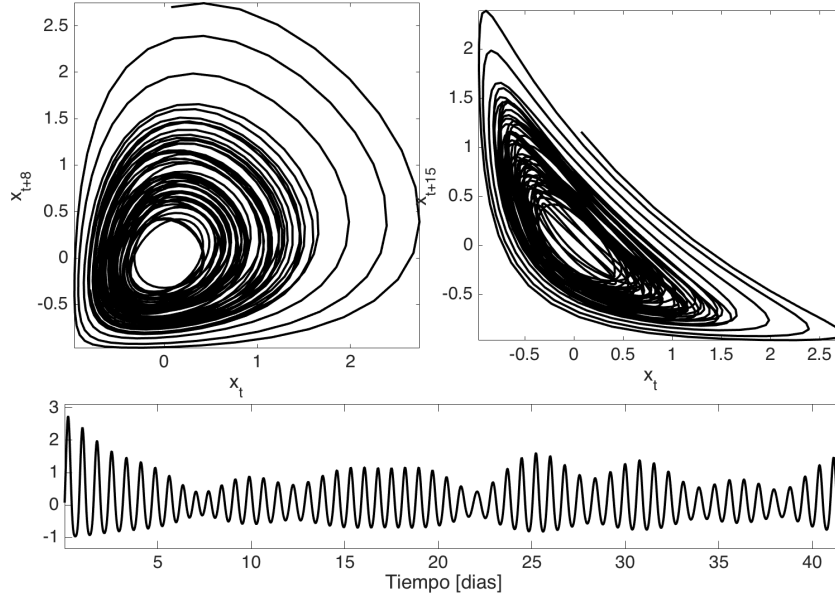


Figura 4.6: Embedding para curva de luz simulada de Gamma Doradus, con amplitudes relativas definidas aleatoriamente. La curva de luz se ilustra en la figura inferior. El espacio de fase de la izquierda se construye con un desfase de $\tau = 8$, mientras que el de la derecha utiliza $\tau = 15$. Se puede observar que un embedding adecuado se puede realizar con tres dimensiones, dado que el atractor no es plano.

En el modelo seleccionado para el análisis se representan ocho frecuencias independientes, aunque muy cercanas entre sí. Dado que el caso ideal es que cada frecuencia sea representada por un oscilador, se prueban modelos que tienen cantidades similares o más osciladores que las frecuencias que se busca representar. Dadas las amplitudes de la tabla 4.5, es esperable que las frecuencias 1.3031, 1.3662 y 1.3997 sean las más difíciles de encontrar. Una correcta identificación de tales frecuencias resultaría una ventaja comparativa frente a un análisis de frecuencia directo de la serie de tiempo.

Como herramienta de comparación, también se realizan pruebas cuando todas las frecuencias del modelo tienen el mismo peso relativo (a_i). Los resultados para los casos de pesos aleatorios y pesos iguales se encuentran descritos en las tablas 4.6 y 4.7 respectivamente. Las pruebas fueron realizadas considerando un tamaño de población de 600 individuos y 300 iteraciones en PSO.

Se puede observar que, en general, a mayor tamaño del modelo se obtienen menores valores de Error Cuadrático Medio, pero tal disminución va de la mano con un aumento en las redundancias o inactividades en los osciladores, además de mayores tiempos requeridos para la optimización debido al número creciente de variables a optimizar. Adicionalmente, los errores de los modelos asociados a la curva de luz simulada con amplitudes aleatorias son siempre menores que aquellos relativos a la señal modelada con amplitudes uniformes. Esto podría ser efecto de la diferencia entre los espacios de búsqueda en ambos casos. Cuando la señal se modela con amplitudes uniformes, todas las componentes de frecuencia aportan de

manera equitativa, y por lo tanto hay una mayor cantidad de óptimos locales de importancia equivalente. Cuando las amplitudes son diferentes, el óptimo local asociado a la frecuencia de mayor amplitud tiene un valor más cercano al óptimo global que un óptimo local asociado a otra de las frecuencias.

Los modelos esféricos presentan ajustes similares o mejores a aquellos obtenidos con modelos de grilla con una cantidad similar de osciladores. Por ejemplo, el modelo esférico de 1 capa consiste de 6 osciladores, menor número que el modelo de grilla de 3x3. En ambos casos para amplitudes relativas iguales y aleatorias, los mejores modelos esféricos obtuvieron un ajuste mínimo mejor que el obtenido con los mejores modelos de 3x3, aunque sus ajustes medio y mediano son peores. Los modelos esféricos de dos capas cuentan con diez osciladores, encontrándose en un número intermedio entre las grillas de 3x3, con nueve osciladores, y las de 3x4, con 12 osciladores. Los valores de mediana y media entre las veinte repeticiones del método con un modelo esférico siguen siendo ligeramente mayores en comparación a aquellos obtenidos con los modelos de grilla.

Tabla 4.6: MSE de los modelos para Gamma Doradus simulada con amplitudes aleatorias.

Modelo	Población	Iteraciones	Mínimo	Mediana	Media
2x2	600	300	0.2314	0.6328	0.5622
3x3	600	300	0.2046	0.2831	0.3511
3x4	600	300	0.1143	0.2480	0.2773
4x4	600	300	0.1096	0.2661	0.2679
5x5	600	300	0.0798	0.2080	0.2122
Esférico, 1 capa	600	300	0.1907	0.4391	0.4668
Esférico, 2 capas	600	300	0.1671	0.3227	0.3759
Esférico, 3 capas	600	300	0.0830	0.2935	0.3223

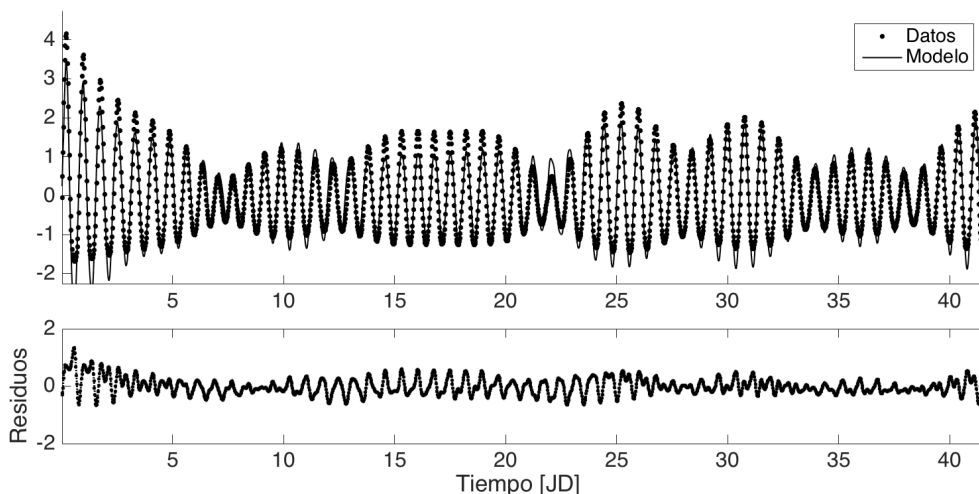


Figura 4.7: Mejor modelo encontrado para la curva de luz de una estrella Gamma Doradus simulada con amplitudes aleatorias. La estructura es una grilla de 5x5. Los residuos aún muestran estructuras que podrían ser simuladas.

El mejor modelo encontrado para señal simulada con amplitudes relativas aleatorias es el de estructura de grilla de 5x5. Su ajuste a la curva y los residuos obtenidos se pueden observar en la figura 4.7. Si bien gran parte de la dinámica es recuperada, el modelo presenta algunas fallas asociadas a sub y sobre ajuste de algunos de los ciclos de pulsaciones. En el caso del mejor modelo para la señal con amplitudes relativas uniformes, el mejor ajuste se obtiene con uno de los modelos de estructura de grilla de 4x4. Al comparar las figuras 4.7 y 4.8 se puede apreciar la diferencia entre los residuos, siendo mayores los errores en el modelo con amplitudes uniformes. Nuevamente, los errores se encuentran asociados a los ciclos de pulsación.

Tabla 4.7: MSE de los modelos para Gamma Doradus simulada con amplitudes uniformes.

Modelo	Población	Iteraciones	Mínimo	Mediana	Media
2x2	600	300	0.4125	0.7309	0.7293
3x3	600	300	0.2746	0.5280	0.5311
3x4	600	300	0.2445	0.5275	0.5491
4x4	600	300	0.1979	0.4427	0.4552
5x5	600	300	0.2052	0.3969	0.4037
Esférico, 1 capa	600	300	0.2672	0.7813	0.7593
Esférico, 2 capas	600	300	0.2259	0.5595	0.5435
Esférico, 3 capas	600	300	0.2820	0.5475	0.5815

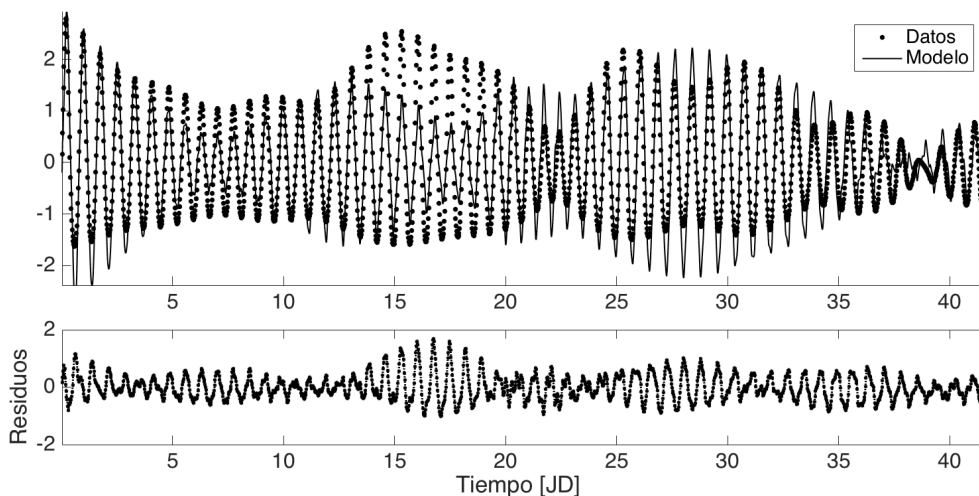


Figura 4.8: Mejor modelo encontrado para la curva de luz de una estrella Gamma Doradus simulada con amplitudes uniformes. La estructura es una grilla de 4x4. Los residuos aún muestran estructuras que podrían ser simuladas, con aspecto pulsante.

4.4. Curvas de luz reales de estrellas Gamma Doradus

4.4.1. Gamma Doradus KIC 9240041

La curva de luz de la estrella Gamma Doradus KIC 9240041 tiene características de modulación de amplitud, aunque sus ciclos no son tan limpios como las señales simuladas descritas en la sección anterior. Cuando la serie de tiempo se representa en un espacio de estados, el diagrama de fase construido en base a la serie trasladada en el tiempo, muestra una forma toroidal, con cierta curvatura apuntando hacia un espacio embebido tri-dimensional. Representaciones bi-dimensionales se ilustran en la figura 4.9, con retrasos temporales de $\tau = 7$ y $\tau = 14$ respectivamente.

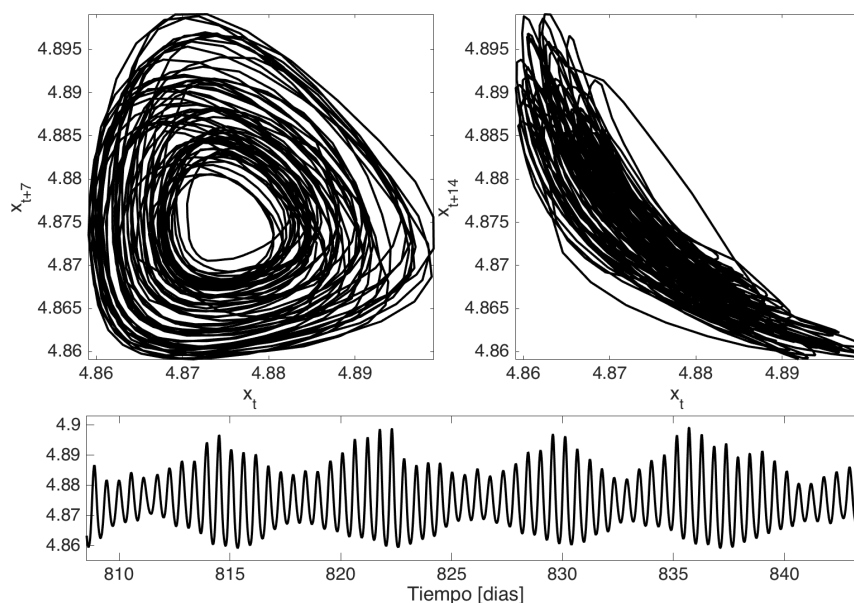


Figura 4.9: Embedding para la curva de luz de la estrella Gamma Doradus KIC 9240041. El gráfico de fase a la izquierda muestra cómo luce el atractor cuando la dimensión del embedding es $M = 2$ y el retardo es $\tau = 7$. El gráfico superior derecho fue creado usando $M = 2$ y $\tau = 14$. Es posible inferir que este atractor tiene una dimensión mayor que 2. La señal original se muestra en la figura inferior.

Dada la naturaleza del proyecto Kepler, esta estrella fue observada de forma extensa. En este trabajo se utilizan únicamente 2000 puntos con observaciones para obtener el modelo de Neuropercolación. El conjunto de datos incluye segmentos con datos perdidos, pero su estructura es aproximadamente equiespaciada en el tiempo. Diferentes estructuras fueron probadas para estos conjuntos de datos, tanto grillas (2×2 , 3×3 , 4×4) como modelos esféricos con una y dos capas. Cada prueba fue repetida veinte veces debido al efecto que tiene la aleatoriedad en los resultados. Los ajustes de MSE para estos modelos se encuentran detallados en la tabla 4.8. Se utilizaron tres configuraciones de Población/Iteraciones para PSO, de manera de analizar la influencia de tales parámetros en los mejores resultados obtenidos.

Tabla 4.8: MSE de los modelos para KIC 9240041

Modelo	Población	Iteraciones	Mínimo	Mediana	Media
2x2	400	400	0.0814	0.1357	0.1354
	600	300	0.0847	0.1702	0.1589
	800	200	0.0842	0.1720	0.1505
3x3	400	400	0.0739	0.0948	0.1119
	600	300	0.0677	0.0902	0.1132
	800	200	0.0622	0.0919	0.1119
4x4	400	400	0.0648	0.0829	0.0935
	600	300	0.0629	0.0816	0.1011
	800	200	0.1020	0.1763	0.1659
Esférico, 1 capa	400	400	0.0839	0.1332	0.1332
	600	300	0.0790	0.1668	0.1343
	800	200	0.0736	0.0983	0.1208
Esférico, 2 capas	400	400	0.0604	0.0888	0.0981
	600	300	0.0674	0.0966	0.1219
	800	200	0.0722	0.0957	0.1178

Los resultados muestran que se obtienen modelos razonablemente buenos para la estrella KIC 9240041 con todas las geometrías probadas. El mejor resultado se obtiene con una la estructura esférica de dos capas, aunque la grilla de 3x3 y la grilla de 4x4 tienen rendimientos similares. Para estructuras más pequeñas se obtienen peores resultados en la modelación, lo que se debe a que en los residuos siguen existiendo estructuras que pueden ser modeladas al añadir una mayor cantidad de osciladores al modelo de Neuropercolación. Pese a ello, un modelo con demasiados osciladores relativo a la cantidad de frecuencias que se quiere representar, puede provocar un sobre-ajuste a los datos. Por otro lado, un mayor número de datos en la curva de luz también puede ayudar a mejorar los modelos. El mejor modelo encontrado se puede observar en la figura 4.10, con sus respectivos residuos.

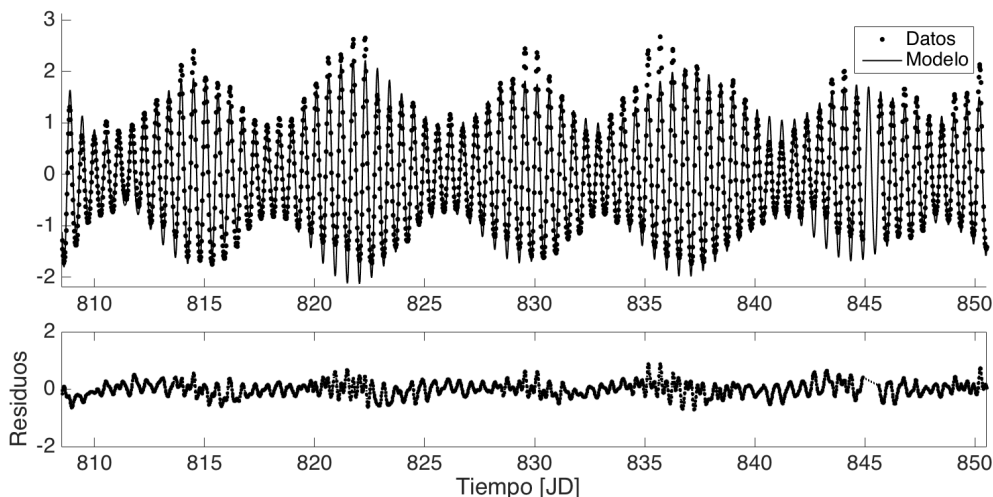


Figura 4.10: Mejor resultado para KIC9240041 obtenido con un modelo esférico de dos capas y sus residuos.

4.4.2. Gamma Doradus KIC 10224094

Esta estrella, también Gamma Doradus, ha sido estudiada por Van Reeth et al. [63]. Aún cuando es parte del mismo tipo de estrellas que la previamente presentada KIC 9240041, al observar la señal y el atractor de la figura 4.11 es claro que no presenta la misma dinámica. Mediante una inspección visual del atractor se puede inferir que su comportamiento es más cercano al caos que a un ciclo límite. Adicionalmente, su composición de frecuencias parece ser mucho más compleja. Según el estudio mencionado, esta estrella es Gamma Doradus con modos progrado de pulsación y sin binariedad, con un período promedio de $0,875860 \pm 0,000110$ días y un espaciamiento de $0,019195 \pm 0,000134$ días entre períodos.

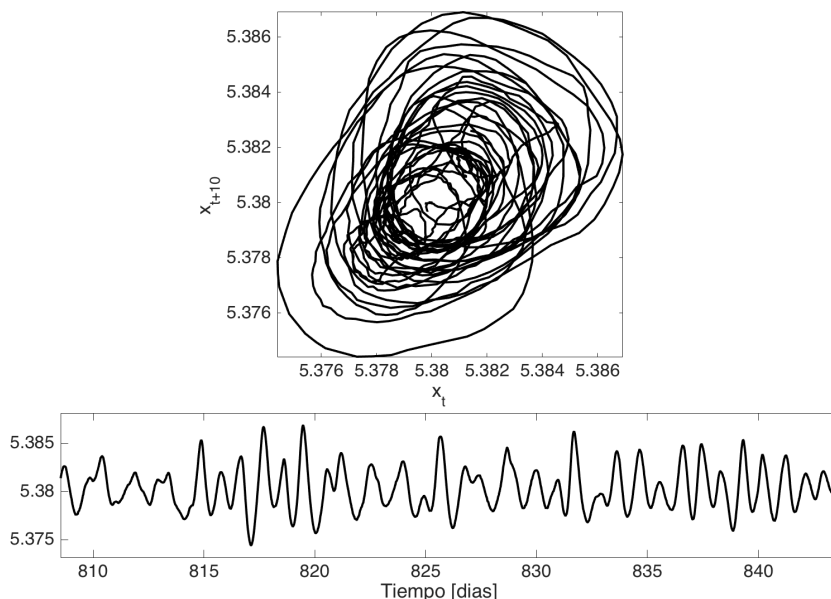


Figura 4.11: Embedding para la estrella Gamma Doradus KIC 10224094. En este caso $M = 2$ y $\tau = 10$. El gráfico inferior corresponde a la serie de tiempo original.

Nuevamente se utilizan únicamente 2000 puntos de la señal para realizar la modelación. El método es probado para diferentes estructuras de Neuropercolación. Se realizan veinte repeticiones de los modelos, con una población de 600 partículas y 300 iteraciones en PSO. Los resultados de éstas pruebas se encuentran en la tabla 4.9, y muestran que esta señal es difícil de modelar.

Tabla 4.9: MSE de los modelos para KIC 10224094.

Modelo	Población	Iteraciones	Mínimo	Mediana	Media
2x2	600	300	0.6515	0.9823	0.9264
2x3	600	300	0.5815	0.7072	0.7444
3x3	600	300	0.5244	0.6595	0.6492
3x4	600	300	0.5284	0.6180	0.6265
4x4	600	300	0.5087	0.6169	0.6220

Aún con el mejor ajuste, presentado en la figura 4.12, se obtienen errores mayores que en el caso precedente. En los residuos ilustrados se observan claramente estructuras oscilatorias que los modelos no logran capturar. La deficiencia en el ajuste puede estar relacionada a diversos efectos, como una elección incorrecta de la combinación población/iteraciones de PSO o una falta de datos en la curva de luz. Otra posibilidad es que las estructuras elegidas no hayan sido adecuadas para representar la dinámica de esta curva de luz.

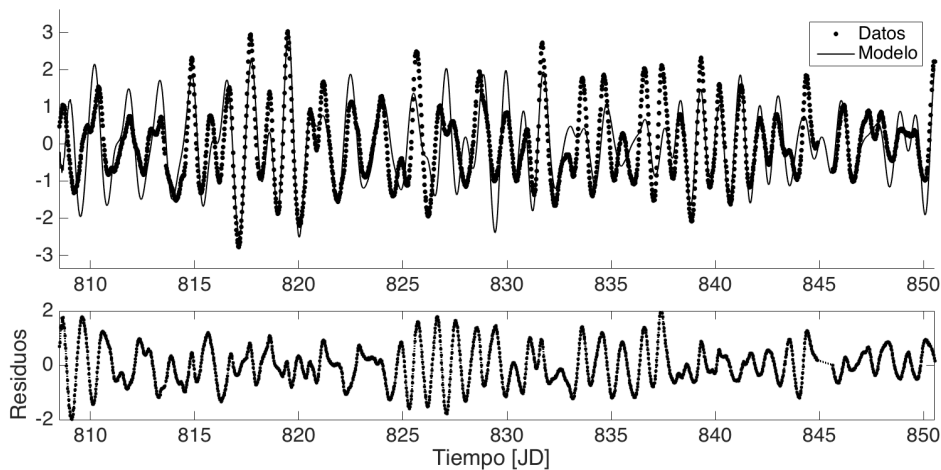


Figura 4.12: Mejor modelo para KIC 10224094 con estructura de grilla de 4x4 y sus residuos. Los residuos muestra estructuras que podrían ser modeladas.

4.4.3. Gamma Doradus KIC 11145123

Esta estrella es un híbrido entre Gamma Doradus y Delta Scuti, un tipo de estrella pulsante con modos de presión. Los modos de presión y gravitacionales se pueden identificar, entre otras cosas, porque se presentan en intervalos distintos de frecuencias. En general las frecuencias de las estrellas Delta Scuti se encuentran entre 5 y 50 días⁻¹ [64], aunque en esta estrella los modos p se encuentran en la banda entre 15 y 25 días⁻¹, mientras que sus modos gravitacionales se encuentran de manera tradicional entre 0.3 y 3 días⁻¹ aproximadamente. Este comportamiento se puede observar directamente en su curva de luz, ilustrada en la figura 4.13, donde se aprecia la dinámica de alta frecuencia montada sobre un comportamiento oscilatorio de frecuencia menor. Adicionalmente, el periodograma de Lomb-Scargle de la señal también ilustra este comportamiento, como se observa en la figura 4.14, donde se diferencian claramente las separación entre los conjuntos de frecuencias.

En trabajos anteriores de Kurtz et al. [65] y Van Reeth et al. [20] se han ajustado modelos de mínimos cuadrados con múltiples frecuencias en tales bandas. En el trabajo de Kurtz et al. adicionalmente se distinguieron los posibles modos gravitacionales asociados a esta estrella. Los períodos de estos modos, en el rango entre 0.5 y 0.8 días presentan un espaciamiento casi uniforme. En tal trabajo se utiliza la información completa de esta estrella obtenida en el proyecto Kepler, y pese que para este trabajo sólo se analizan 2000 puntos de la curva de luz, teóricamente es posible identificar tales períodos, siempre y cuando sus modos de oscilación hayan sido excitados en el momento de observación analizado.

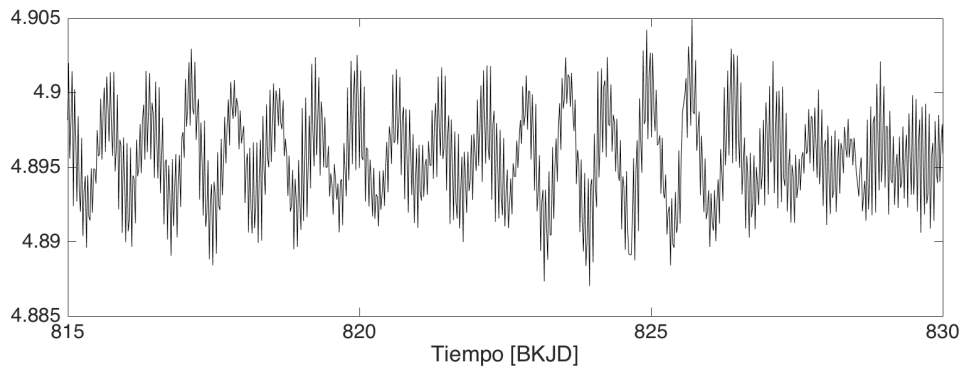


Figura 4.13: Parte de la curva de luz utilizada de la estrella KIC 11145123, un híbrido entre Gamma Doradus y Delta Scuti. Esta estrella presenta modos de presión, observables como pulsaciones de mayor frecuencia, y modos gravitacionales de oscilación.

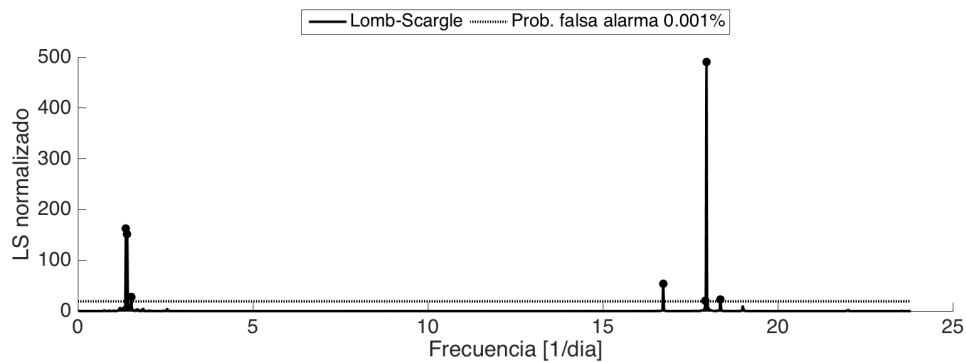


Figura 4.14: Periodograma de Lomb Scargle de la curva de luz de la estrella KIC 11145123. Se observan dos conjuntos de frecuencias principales. Las frecuencias altas se encuentran asociadas a los modos de pulsación de presión de la estrella, mientras que las frecuencias entre 0.3 y 3 días son las producidas por los modos gravitacionales.

Se prueban estructuras similares a las ya analizadas para otras curvas de luz, realizando veinte repeticiones de PSO para cada una, con un tamaño de población de 600 individuos y 300 generaciones. Los resultados de estos modelos se encuentran detallados en la tabla 4.10, donde se observa que los ajustes de la curva de luz no son tan exactos como aquellos encontrados para KIC 9240031, aunque son mejores que aquellos de KIC 10224094. Los valores mínimos de MSE son similares para todas las estructuras, un efecto ya observado en la curva de luz simulada de una estrella Gamma Doradus con amplitudes relativas uniformes. Nuevamente, esto puede interpretarse como un efecto de la forma del espacio de búsqueda, en lo cual se ahondará en la sección de Análisis.

El mejor resultado corresponde a un modelo de grilla de 3x3, el cual se encuentra graficado en la figura 4.15 junto con sus residuos. Tales residuos presentan componentes de alta y baja frecuencia, lo cual puede ser deducido por la estructura que presenta. En la figura 4.16 se grafica el periodograma de Lomb-Scargle de los residuos, donde se aprecia que la principal frecuencia que no se modela bien es alta, asociada a los modos de presión. Esta frecuencia corresponde al segundo máximo del conjunto de altas frecuencias de la señal original.

Tabla 4.10: MSE de los modelos para KIC 11145123

Modelo	Población	Iteraciones	Mínimo	Mediana	Media
2x2	600	300	0.2797	0.4894	0.6819
3x3	600	300	0.2273	0.3853	0.4863
3x4	600	300	0.2500	0.3705	0.4038
4x4	600	300	0.2359	0.3535	0.4058
Esférico, 1 capa	600	300	0.2538	0.4585	0.6213
Esférico, 2 capas	600	300	0.2472	0.3482	0.4241

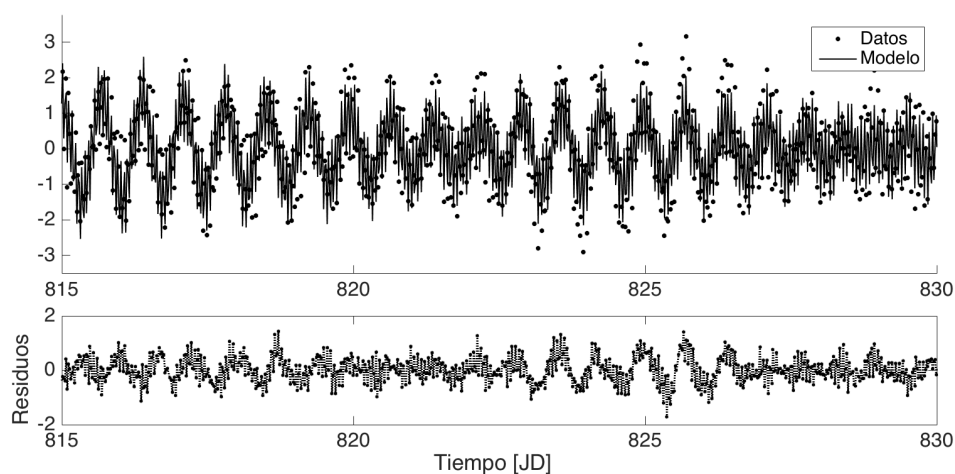


Figura 4.15: Detalle del ajuste del mejor modelo para KIC 11145123, una grilla de 3x3, y sus residuos. En los residuos siguen presentes componentes de alta y baja frecuencia.

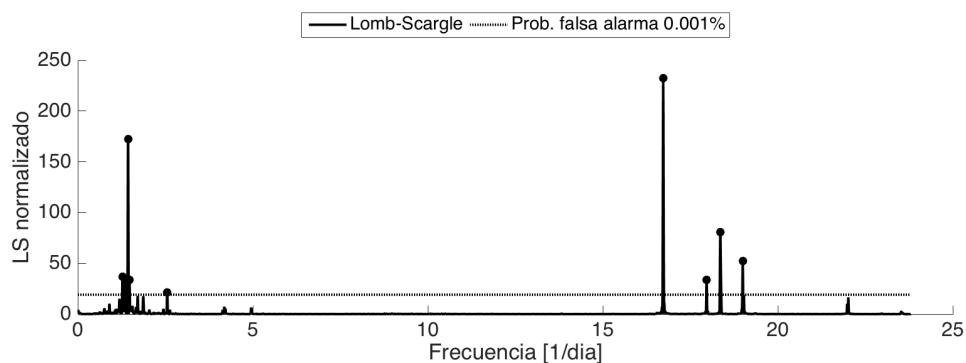


Figura 4.16: Periodograma de Lomb-Scargle de los residuos del mejor modelo para KIC 11145123. Se observa que quedan componentes de alta y baja frecuencia por modelar.

5 Análisis y discusión de resultados

5.1. Señales Sintéticas

5.1.1. Señales de Amplitud Modulada

Dado el modelo de amplitud modulada presentado en el capítulo anterior, se espera encontrar osciladores asociados a la suma y diferencia de las frecuencias de la señal de información y de la señal portadora. Ambas frecuencias fueron encontradas al analizar los mejores resultados obtenidos mediante modelos Neuropercolación. En los casos donde los modelos no alcanzaron un buen ajuste, éstos representaban únicamente una de las dos frecuencias buscadas. Se puede asumir de manera razonable que estos modelos se deben encontrar cercanos a mínimos locales en el espacio de características. Un ejemplo de esto es el peor resultado obtenido al utilizar el 10 % de los datos, que se encuentra graficado en la figura 5.1. En este, al igual que en otros casos sólo se identifica una de las frecuencias asociadas a la modulación de amplitud.

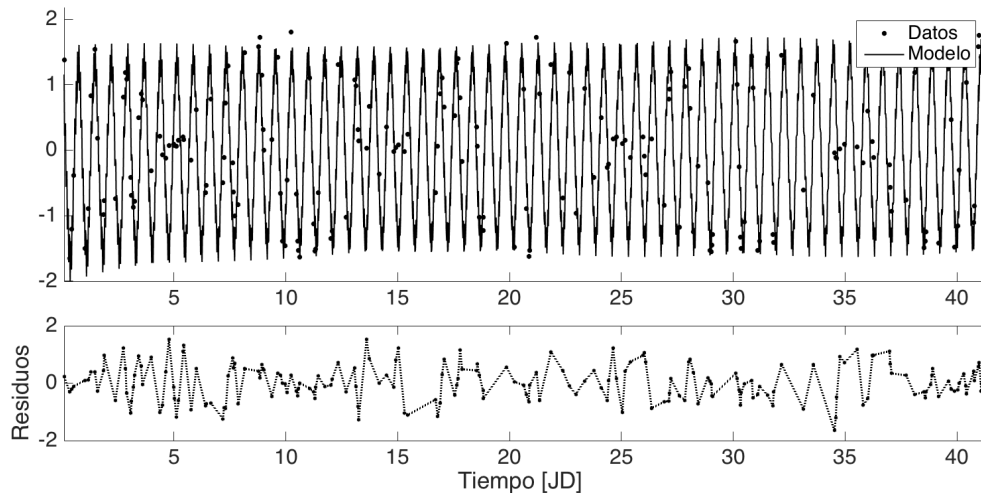


Figura 5.1: Peor modelo encontrado para señal AM con el 10 % de los datos. El modelo se ajusta a sólo una de las dos frecuencias que se desea encontrar.

Al utilizar información incompleta para realizar la modelación se observa que sigue siendo posible obtener modelos que representaran de manera confiable la dinámica de la señal. Al realizar veinte repeticiones, algunas de éstas no llegan a un resultado favorable, pero en todas las pruebas el mejor caso presenta un ajuste de mínimos cuadrados de orden similar. Aparentemente se pierde uniformidad en los resultados al disminuir la cantidad de información disponible, lo que se traduce en un peor ajuste medio y de mediana en los casos con más datos perdidos. Dados los resultados, se podría estudiar si un mayor número de datos en la serie de tiempo producen resultados más uniformes, disminuyendo la cantidad de modelos que convergen a óptimos locales.

Para los análisis de frecuencia, tanto sobre la señal original como sobre la descomposición de los modelos de Neuropercolación se utiliza el periodograma de Lomb-Scargle. Para identificar las frecuencias máximas se busca simplemente el máximo global del periodograma. Para identificar todas las componentes de frecuencia se buscan los máximos locales, descartando aquellos bajo el umbral de 0.001 % de falsa alarma.

Al analizar las frecuencias que toman los distintos osciladores se observa que se representan las dos frecuencias relevantes esperadas, que son la suma y resta de las frecuencias del modelo de señal de Amplitud Modulada. Los mejores resultados, cuyas frecuencias se detallan en la tabla 4.2, siempre representan estas dos frecuencias, incluso con sólo el 10 % de los datos. Los cuatro osciladores otorgan flexibilidad, lo que se ve al analizar las frecuencias de cada oscilador. En algunos casos se repiten frecuencias, o hay osciladores que se mantienen inactivos.

Es claro que los modelos encontrados representan de buena manera la señal, tanto en tiempo como en frecuencia, pero también se debe analizar si es que esto entrega alguna ventaja comparativa. Cuando se analizan los osciladores por separado, las frecuencias que son máximo global son, en general, aquellas que se desea encontrar; pero hay otras frecuencias que son máximos locales en los periodogramas de cada oscilador. En la figura 5.2 se grafica el periodograma de Lomb Scargle de la señal, las frecuencias objetivo y las frecuencias principales de todos los osciladores, tanto los máximos globales como locales de sus periodogramas¹. En esta figura se aprecia que hay frecuencias adicionales a las buscadas, relacionadas a la utilización de Lomb-Scargle en señales con ventanas de tiempo son rectangulares. Este efecto se puede reducir mediante la utilización de ventanas sobre los datos, o mediante el aumento de datos. Estas mismas frecuencias también son máximos locales en el periodograma aplicado a la señal original, pero la descomposición de señales produce que sus amplitudes se vean aumentadas. Esta es una propiedad que se busca aprovechar para encontrar componentes de frecuencia que puedan confundirse con ruido al estudiar la composición de frecuencias de la señal. Al disminuir la cantidad de datos, los mejores modelos muestran una identificación consistente de las frecuencias buscadas, las cuales destacan cada vez más por sobre un periodograma de la señal que presenta cada vez menores valores.

¹En el Anexo 7.3 se explica en mayor detalle la forma de obtención de las frecuencias asociadas a los osciladores en los modelos de Neuropercolación.

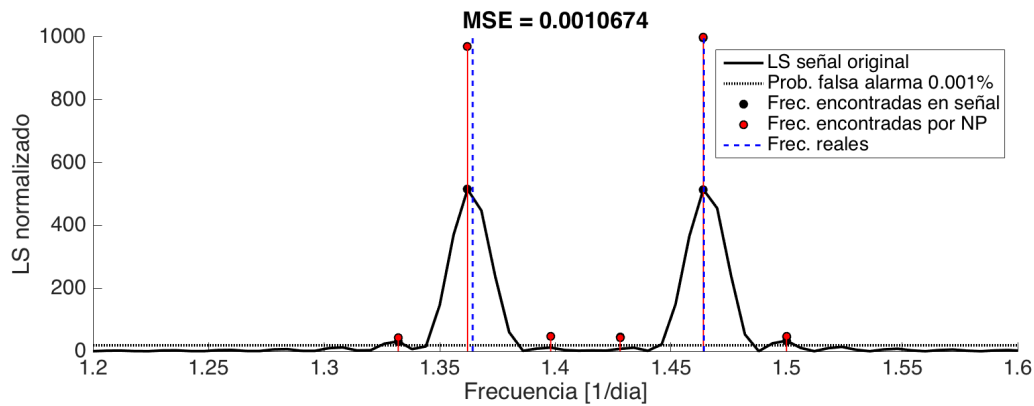


Figura 5.2: Frecuencias encontradas para todos los osciladores en el mejor modelo de señal AM sin datos perdidos. Se escogen las frecuencias que son máximos locales y sobrepasan el límite de 0.001% de falsa alarma en el periodograma de Lomb Scargle.

Cuando se añade ruido blanco aditivo los resultados muestran que es posible reconstruir la señal, a pesar de altos niveles de ruido. Esto se encuentra relacionado al modelo seleccionado, dado que limita la cantidad de frecuencias que se pueden representar de forma precisa. Este efecto puede considerarse como un beneficio o como una limitación del modelo. Se debe tener una idea previa de cuantas frecuencias debe poseer la dinámica principal, de modo de escoger un modelo adecuado. Si se escogiera un modelo con mayor capacidad de representación, se produciría un sobre ajuste al ruido aditivo, provocando el aumento de componentes de alta frecuencia.

Como es de esperar, los valores de MSE entre los modelos de Neuropercolación y las señales ruidosas aumentan con el nivel de ruido. Pese a ello, al comparar los modelos con la señal original sin ruido, se obtuvo que los valores de MSE se encuentran en rangos similares, independiente del nivel de ruido. Esto lleva a concluir que los modelos no se sobre-ajustan al ruido blanco aditivo. Es especialmente importante que los valores mínimos se encuentran en rangos similares. Los MSE de los mejores modelos con SNR de 10 y 5 presentan los valores más altos del conjunto, incluso mayores que con SNR igual a 1, lo cual puede ser atribuido a la aleatoriedad intrínseca de PSO. Adicionalmente, estos mismos conjuntos de resultados (SNR 10 y 5) son los que presentan valores mayores en sus medianas y medias, probablemente por el efecto de las poblaciones iniciales.

Otra forma de reforzar la idea de que los modelos no se ajustan al ruido aditivo es el observar la distribución de los errores de modelación. En la figura 5.3 se observan los histogramas de los errores de modelación ($y_{datos} - y_{modelo}$) para los mejores modelos en todos los niveles de ruido. Se observa que los errores distribuyen con forma gaussiana, siendo su varianza superior para los niveles de mayor ruido. Adicionalmente, las figuras muestran la distribución teórica que debería tener el error entre la señal original y las señales para diferentes niveles de ruido. Se observa que las distribuciones empíricas encontradas con los mejores modelos son muy similares a las teóricas. Esto significa que el error de modelación está asociado casi únicamente al ruido de las señales con las que se entrenan los modelos.

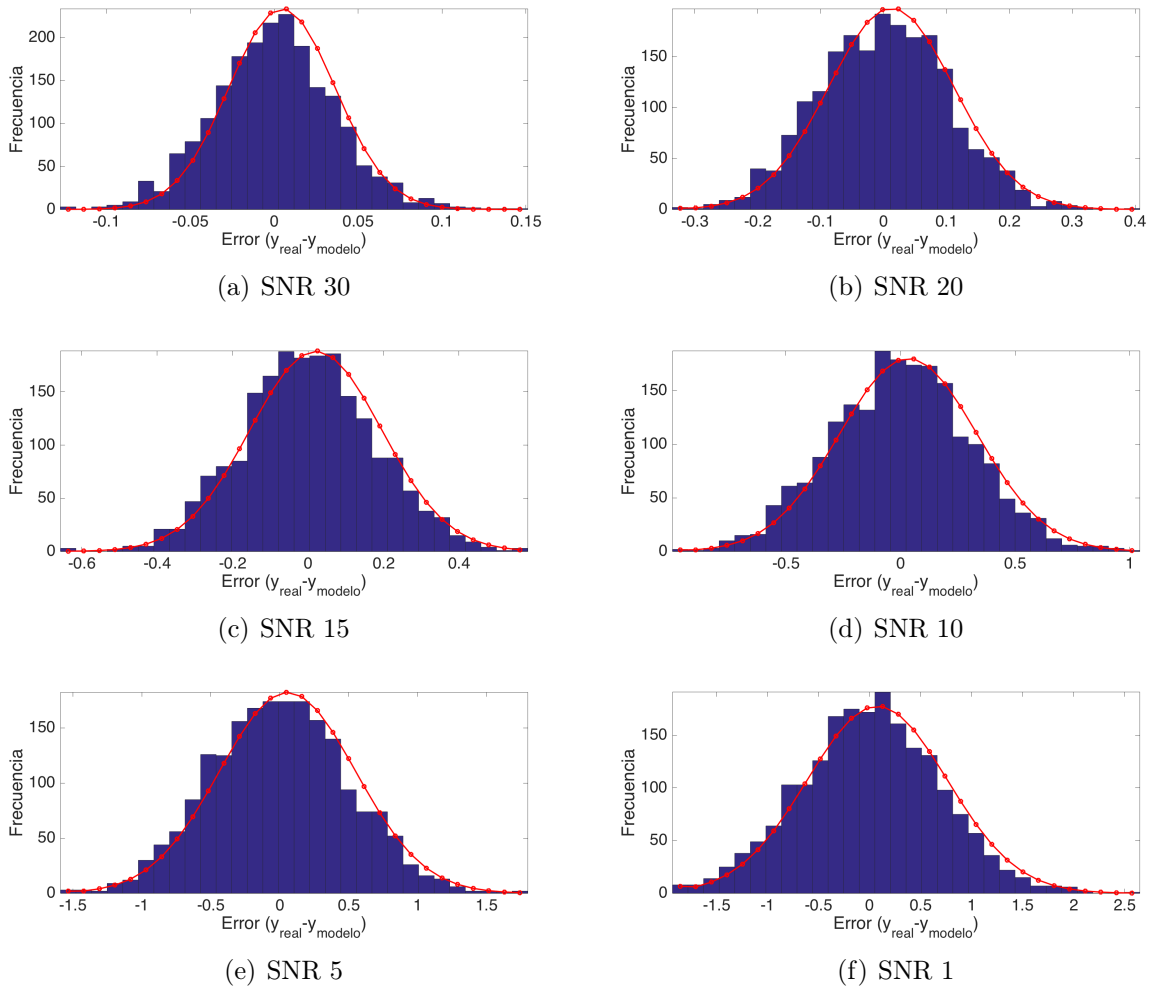


Figura 5.3: Histogramas de error en la modelación de señales con ruido aditivo. Se observa que las distribuciones tienen forma gaussiana y que su varianza aumenta con el nivel de ruido. La línea roja corresponde a la distribución teórica del error asociado al ruido. La similitud entre las distribuciones empíricas y teóricas señala que los modelos no se sobre-ajustan al ruido.

De manera similar a lo que ocurre al disminuir la cantidad de datos, el periodograma de la señal con ruido disminuye su potencia a medida que el SNR disminuye. Para los mejores modelos obtenidos la identificación de las frecuencias relevantes se mantiene, existiendo aún máximos locales en frecuencias adicionales. Esto se puede observar para el caso de SNR 30 en la figura 5.4.

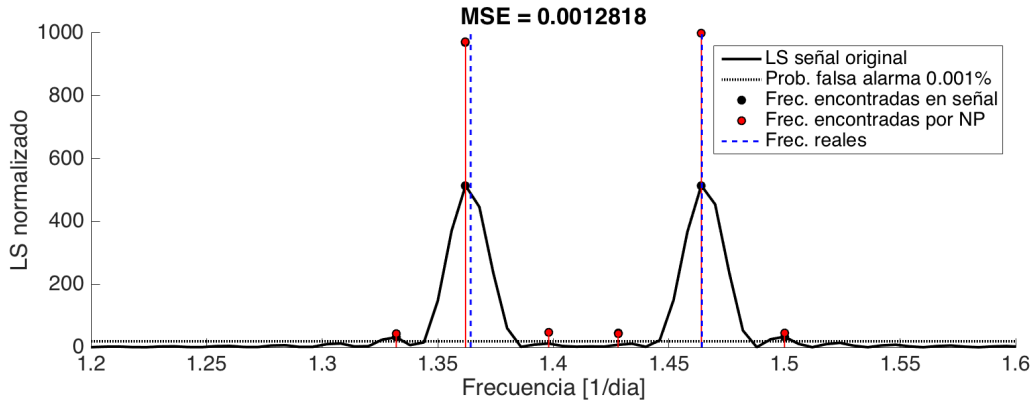


Figura 5.4: Frecuencias encontradas para todos los osciladores en el mejor modelo de señal AM con SNR 30. Se escogen las frecuencias que son máximos locales y sobrepasan el límite de 0.001 % de falsa alarma en el periodograma de Lomb Scargle.

Al analizar cada oscilador de manera separada se puede observar que algunos no oscilan, tal y como se aprecia en las tablas 4.2 y 4.4. En el primer caso hay osciladores que se encuentran inactivos para los mejores modelos utilizando 100, 80, 60 y 40 % de los datos. En la segunda tabla se observa que lo mismo ocurre para los mejores modelos obtenidos con señales con SNR de 10 y 20. Pese a ello, incluso cuando un oscilador particular no presenta variaciones, puede tener un rol relevante dependiendo de las conexiones que tenga con sus vecinos. Por ejemplo, puede tener un efecto regulatorio en la fase de otra señal.

En el caso de que uno de los conjuntos KII no presente variaciones y sus conexiones con otros osciladores sean nulas, éste puede ser removido para crear un modelo más compacto. Este es el caso del mejor modelo obtenido utilizando el 60 % de los datos, que es el ilustrado en la figura 5.5(a). En este caso el oscilador inferior izquierdo tiene como parámetros $p = 0$ y $\omega = 1$, los cuales son una combinación que lleva a una dinámica de punto fijo, es decir, no presenta oscilaciones. Adicionalmente sus conexiones son nulas, por lo que no interactúa con el resto de los conjuntos KII de la grilla. Por ese motivo es posible eliminarlo, construyendo una grilla de 1×3 , como se observa en la figura 5.5(b). En el mismo ejemplo, otros osciladores también se encuentran aislados del resto, presentando conexiones nulas con los demás osciladores. Pero si tales osciladores presenten dinámicas diferentes al punto fijo, el eliminarlos provocaría un cambio en la dinámica del modelo de Neuropercolación.

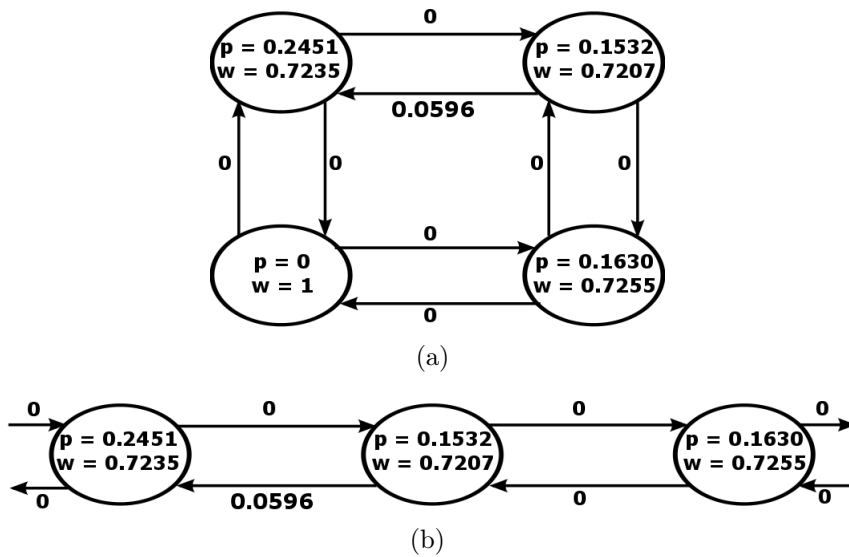


Figura 5.5: Modelos equivalentes para una señal de Amplitud Modulada. En el modelo 2x2 de la figura 5.5(a) el oscilador inferior izquierdo se encuentra inactivo y desconectado del resto. Esto se puede transformar en un modelo de 1x3, eliminando tal oscilador, como se muestra en la figura 5.5(b)

5.1.2. Curva de luz simulada de estrella Gamma Doradus

El objetivo principal de analizar esta curva de luz simulada es poder reconstruir la base de períodos/frecuencias que la componen. Como ya fue mencionado, la principal dificultad consiste en que no todas las componentes aportan de manera equitativa, sino que hay algunas con amplitudes relativas muy pequeñas. Se analiza principalmente la descomposición en osciladores y las frecuencias que éstos logran recuperar, lo cual se encuentra relacionado con la capacidad de reconstrucción del espaciamiento entre períodos, pieza clave del estudio de las estrellas Gamma Doradus.

Tal como en el caso anterior, para realizar un análisis en frecuencia se estudian los osciladores de los modelos de Neuropercolación de forma separada. Se calcula el periodograma de Lomb-Scargle de cada oscilador, identificando sus máximos locales. Si estos máximos superan el umbral de falsa alarma de 0.001 %, se escogen como componentes de frecuencia del modelo.

En la figura 5.6 se presenta el mejor resultado para una topología cuadrada de 4x4. Se destacan todas las frecuencias asociadas a los osciladores, tanto los máximos globales como los locales, siempre y cuando sean mayores al umbral definido por la probabilidad de falsa alarma. Se observa que las frecuencias cuyas amplitudes relativas son mucho menores al resto (1.3031, 1.3662 y 1.3997 días⁻¹), no son identificadas como principales en los osciladores. Pese a ello, sí son identificadas como máximos locales, por lo menos en el entorno. En cambio, el periodograma de Lomb-Scargle sobre la señal original falla en identificar las frecuencias de baja amplitud, sin presentar máximos cercanos a 1.3031 y 1.3662 días⁻¹. Pese a que el modelo con menor MSE corresponde a una estructura de 5x5, el modelo presentado de 4x4 identifica de mejor manera las frecuencias de la curva.

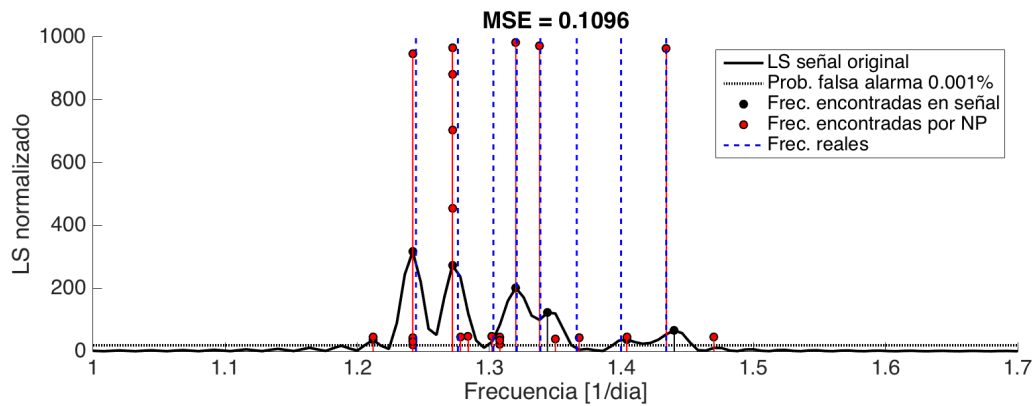


Figura 5.6: Frecuencias encontradas para curva de luz simulada con modelo de 4x4. La línea negra corresponde al periodograma de Lomb-Scargle sobre la señal. Las líneas punteadas azules son las frecuencias usadas para generar la señal. Los puntos rojos son los máximos de los espectros de todos los osciladores del modelo de Neuropercolación. El periodograma de Lomb-Scargle no presenta máximos locales en dos de las frecuencias de la señal. El modelo de Neuropercolación presenta máximos locales en todas las frecuencias, aunque la potencia asociada a algunas de éstas es baja.

Para analizar el espaciamiento, primero hay que limpiar la información cruda sobre frecuencias encontradas. Se realiza un proceso de eliminación simple, que consiste en los siguientes pasos:

- Transformar las frecuencias a períodos.
- Identificar los máximos locales del periodograma.
- Seleccionar los máximos que sobrepasen el límite definido por una probabilidad de falsa alarma de 0.001 %.
- Seleccionar sólo los períodos que se encuentran entre 0.3 y 3 días, que es el rango esperado para estrellas Gamma Doradus.
- Si un par de períodos presentan una diferencia porcentual de menos de 1 % en sus valores, con respecto al período menor, se mantiene solamente aquel con un valor mayor en el periodograma.

Este proceso asegura que no haya períodos similares que puedan afectar el cálculo del espaciamiento. El mejor resultado de un modelo plano de Neuropercolación de tamaño 4x4 presenta el espaciamiento ilustrado en la figura 5.7. Se observa que debido a que Lomb-Scargle sobre la señal original no identifica todos los períodos relevantes, el espaciamiento logrado es lejano al ideal. El modelo de Neuropercolación, con los períodos seleccionados de la manera explicada anteriormente se asemeja mucho al espaciamiento teórico. Se observa que un período adicional es añadido, además de que errores en la estimación de las frecuencias provocan diferencias en los espaciamientos.

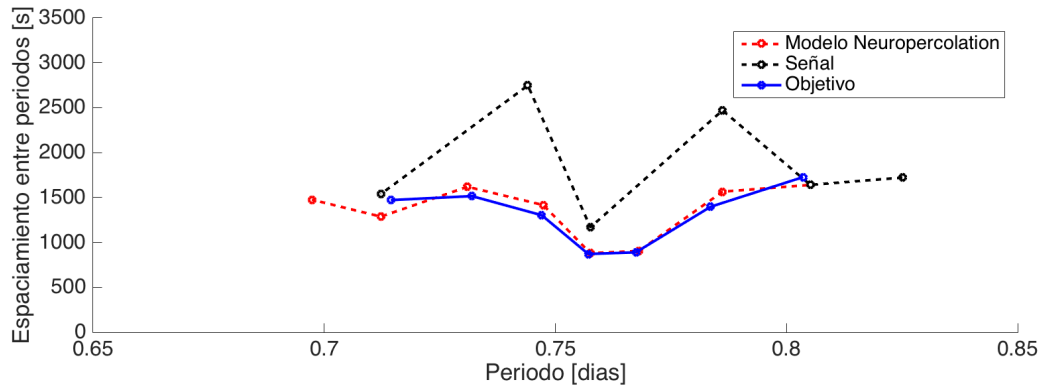


Figura 5.7: Espaciamiento de períodos para curva de luz simulada de amplitudes aleatorias de una estrella Gamma Doradus con el mejor modelo de Neuropercolación encontrado de tamaño 4x4.

Para los mejores modelos encontrados para cada estructura, es posible analizar la capacidad de recuperación de los períodos/frecuencias, de manera independiente a su bondad de ajuste. Se puede observar en la tabla 5.1 para el modelo de amplitudes aleatorias, y en la tabla 5.2 para el modelo de amplitudes uniformes, cuáles frecuencias son identificadas por los mejores modelos. Se acepta una identificación de frecuencia si tiene un error porcentual máximo de 0.5 % con respecto a la frecuencia real.

Tanto en el caso de la curva simulada con amplitudes relativas aleatorias, como aquel con amplitudes relativas uniformes, se observa que la mayor capacidad de representación de frecuencias se alcanza con modelos de estructura de grilla de 4x4, que contienen 16 osciladores. Cuando se tienen menos osciladores se pierden ciertas frecuencias, probablemente por falta de capacidad de estimación, pero si se aumenta demasiado la cantidad de osciladores algunos pueden quedar inactivos o ser redundantes, además de aumentar considerablemente el tamaño del espacio de búsqueda. Esto ocurre, por ejemplo, al utilizar estructuras de 5x5, donde a pesar de encontrar un mejor ajuste en el tiempo, se pierde calidad en la estimación de frecuencias. Cabe destacar que a pesar de que un modelo logre recuperar todas las frecuencias, no necesariamente se va a poder construir un espaciamiento adecuado, dado que pueden existir frecuencias adicionales en el modelo. Este es el caso del modelo esférico de tres capas para la señal con amplitudes aleatorias. La presencia de frecuencias adicionales puede estar relacionada con un sobre-ajuste en el modelo.

Tabla 5.1: Frecuencias identificadas por los mejores modelos de Neuropercolación para una curva de luz simulada de una estrella Gamma Doradus con ocho frecuencias de amplitudes aleatorias.

Frecuencia	2x2	3x3	3x4	4x4	5x5	Esf. 1	Esf. 2	Esf. 3	L-S
1.2444	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.2762	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.3031	✓	✓	✓	✓	✓	✓	✓	✓	
1.3208				✓	✓	✓	✓	✓	✓
1.3386		✓	✓	✓	✓		✓	✓	✓
1.3662		✓		✓			✓	✓	
1.3997				✓	✓			✓	✓
1.4339			✓	✓	✓	✓		✓	✓

Tabla 5.2: Frecuencias identificadas por los mejores modelos de Neuropercolación para una curva de luz simulada de una estrella Gamma Doradus con ocho frecuencias de amplitudes uniformes.

Frecuencia	2x2	3x3	3x4	4x4	5x5	Esf. 1	Esf. 2	Esf. 3	L-S
1.2444	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.2762	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.3031	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.3208				✓	✓				
1.3386	✓	✓	✓	✓	✓	✓	✓	✓	
1.3662	✓	✓	✓					✓	
1.3997	✓		✓	✓	✓	✓	✓	✓	✓
1.4339				✓	✓	✓			✓

5.2. Curvas de luz reales de estrellas Gamma Doradus

5.2.1. Gamma Doradus KIC 9240041

Esta estrella presenta un comportamiento bastante marcado de cuasi-periodicidad, observable tanto en su curva de luz como en su representación de espacio de estado. En esta última representación luce muy similar al modelo simulado de Gamma Doradus utilizado previamente. Por ese motivo es un buen punto de comparación para analizar el comportamiento de la metodología en casos reales. Desafortunadamente, no existe información sobre sus espaciamentos de períodos, por lo que estos resultados no pueden ser comparados directamente.

En cuanto a la representación en frecuencia de la señal original, se observa que el periodograma de Lomb-Scargle presenta un único máximo de mayor relevancia, con algunos máximos locales. En cambio, al realizar un análisis de las frecuencias que aparecen en cada uno de los osciladores del mejor modelo con estructura 4x4 se observa que hay múltiples frecuencias encontradas. Ambas características se encuentran graficadas en la figura 5.8.

Algunas de las frecuencias presentes en los osciladores se observan también como máximos locales en el periodograma de Lomb-Scargle de la señal, pero otras no se observan en éste. Utilizando las frecuencias encontradas en el modelo se puede construir un espaciamiento entre períodos, con la misma metodología planteada para el caso de las curvas de luz simuladas. El resultado se muestra en la figura 5.9. Existen grandes diferencias de espaciamiento al analizar los osciladores por separado y al estudiar directamente la señal utilizando Lomb-Scargle. En el primer caso se obtienen períodos menores y espaciamientos más uniformes, que asemejan la forma teórica esperada. Aún así, estos períodos y sus correspondientes espaciamientos no han sido identificados por astrónomos, por lo que no pueden ser calificados como certeros.

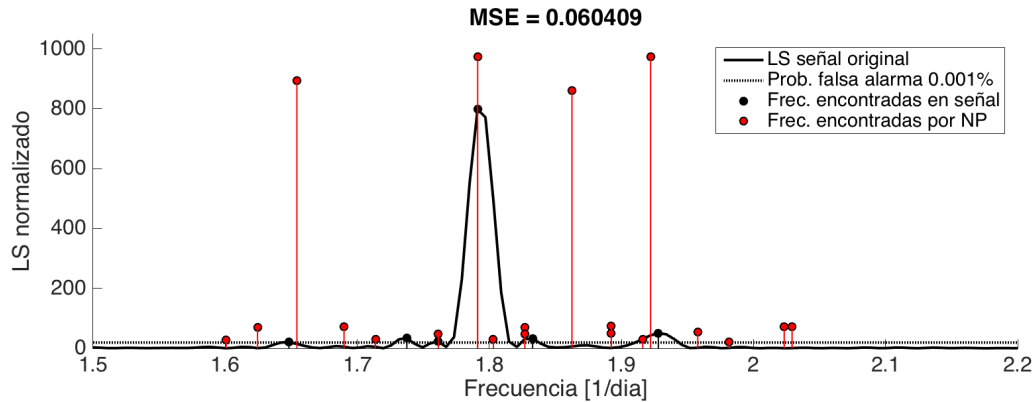


Figura 5.8: Frecuencias encontradas para curva de luz KIC 9240041 con el mejor modelo esférico de dos capas. La curva negra corresponde al periodograma de Lomb-Scargle de la señal original. Los puntos rojos corresponden a las frecuencias encontradas al analizar los osciladores del modelo de Neuropercolación.

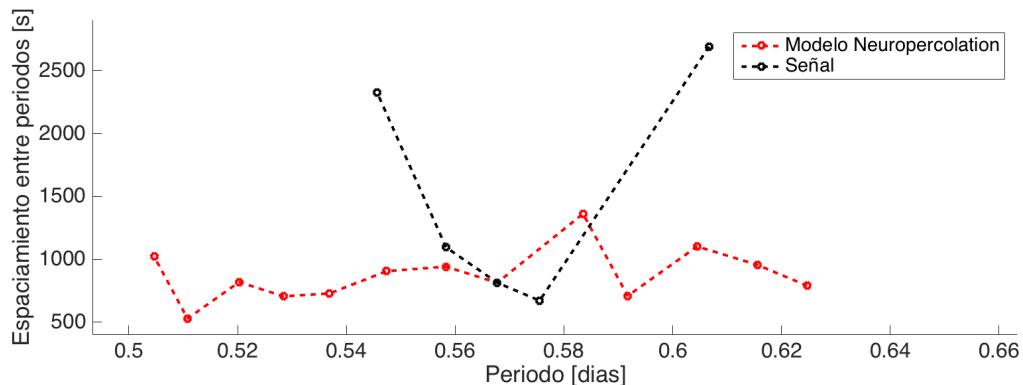


Figura 5.9: Espaciamiento de períodos para KIC9240041 construido con el mejor modelo encontrado de tamaño esférico de dos capas.

La curva de luz de esta estrella muestra un gran potencial para la modelación, siendo la curva de luz real con la que se logró un mejor ajuste. Incluso con modelos pequeños, como grillas de 2×2 , se llega a valores bajos de MSE. Esta característica podría ser explotada en el estudio de cómo realizar transferencia de información entre modelos de Neuropercolación para optimizar modelos de mayor envergadura con una inicialización de la población que incluya información *a priori*.

Adicionalmente, la utilización de una mayor cantidad de datos podría mejorar la calidad del modelo y hacerlo más general. En este trabajo se analizaron únicamente 2000 puntos de la curva de luz, que pueden aumentarse considerablemente. Se analizó el MSE en los siguientes 2000 puntos de la curva de luz para todos los modelos encontrados mediante el proceso de optimización, obteniendo los resultados presentados en la tabla 5.3. Si estos resultados se comparan con la tabla 4.8 es claro que el ajuste empeora en los datos de prueba. Pese a ello, algunos de los modelos logran un ajuste bastante bueno, aunque inferior al logrado con los datos de entrenamiento.

Tabla 5.3: MSE con datos de prueba en los modelos encontrados para KIC 9240041.

Modelo	Población	Iteraciones	Mínimo	Mediana	Media
2x2	400	400	0.1092	0.1762	0.3172
	600	300	0.0932	0.1879	0.1798
	800	200	0.1132	0.1864	0.5565
3x3	400	400	0.0947	0.1207	0.1369
	600	300	0.0980	0.1556	4.3063
	800	200	0.0946	0.1308	0.1449
4x4	400	400	0.0832	0.1186	0.2964
	600	300	0.0914	0.1198	0.5460
	800	200	0.1564	0.2012	0.2768
Esférico, 1 capa	400	400	0.0967	0.1848	0.1570
	600	300	0.1061	0.1858	9.5718
	800	200	0.0996	0.1776	0.1597
Esférico, 2 capas	400	400	0.0796	0.1132	0.1203
	600	300	0.0901	0.1477	0.3248
	800	200	0.0935	0.1310	14.6169

5.2.2. Gamma Doradus KIC 10224094

La curva de luz de esta estrella es la que tuvo peor rendimiento en modelación, por lo que los resultados obtenidos no representan de forma exacta la señal. Aún así, se puede obtener cierta información mediante el mejor modelo encontrado.

En la figura 5.10 se observan las frecuencias encontradas para el mejor modelo, correspondiente a una estructura de grilla de 4x4. Los espaciamientos asociados a este mismo modelo se encuentran graficados en la figura 5.11, comparándolos con aquellos obtenidos mediante el uso directo del periodograma de Lomb-Scargle en la curva de luz. Según Van Reeth et al. [63] el espaciamiento promedio entre períodos es de $0,019195 \pm 0,000134$ días, o $1197,8 \pm 8,3616$ segundos. Si bien los espaciamientos tanto del modelo como de la señal se encuentran lejanos a tales valores, aquellos obtenidos desde el modelo de Neuropercolación son generalmente de menor orden que los de la señal, acercándose más a los valores reales. No obstante, es posible asumir, dada la información obtenida de la literatura, que muchas componentes de frecuencia no se encuentran representadas en ninguno de los dos espectros presentados.

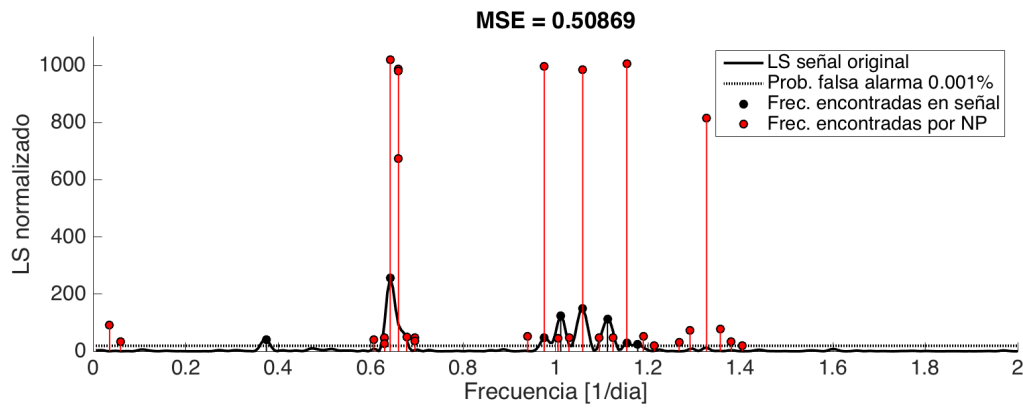


Figura 5.10: Frecuencias encontradas para curva de luz KIC 10224094 con el mejor modelo de 4x4. La línea negra corresponde al periodograma de Lomb-Scargle sobre la señal original.

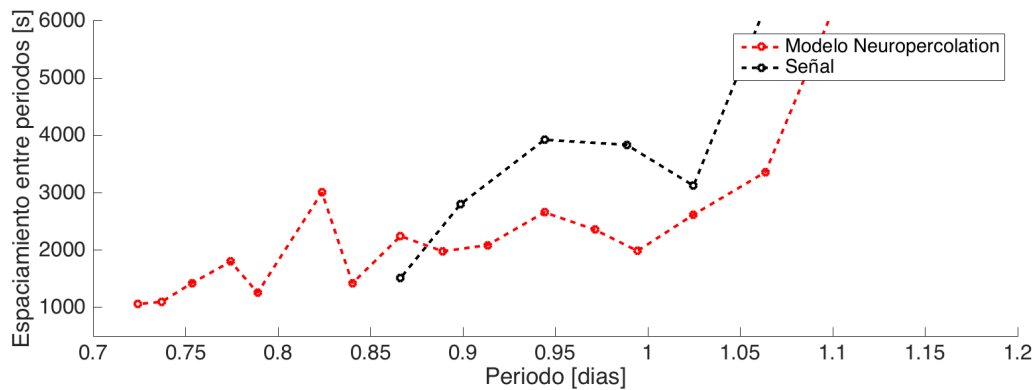


Figura 5.11: Espaciamientos encontrados para curva de luz KIC 10224094 con el mejor modelo de 4x4.

5.2.3. Gamma Doradus KIC 11145123

Al analizar los modelos obtenidos para esta estrella híbrida se puede apreciar que se emulan las dinámicas más notorias: la alta frecuencia asociada al comportamiento de Delta Scuti y un par de frecuencias más bajas, asociadas al comportamiento de Gamma Doradus de la estrella. Sin embargo, ningún modelo obtiene un ajuste sobresaliente. Más aún, para todas las estructuras probadas los mejores modelos tienen el mismo orden de ajuste en mínimos cuadrados. De esto se puede inferir que probablemente logran simular las mismas componentes de frecuencia. Esta hipótesis se puede reafirmar al observar la tabla 5.4. Esta tabla muestra los períodos asociados a los modos gravitacionales identificados por Kurtz et al. [65]. Se observa que, al igual que el periodograma de Lomb-Scargle utilizado directamente sobre la señal, los mejores modelos identifican dos frecuencias principales. Sólo los mejores modelos encontrados para estructuras de 3x3 y 3x4 identifican un tercer periodo.

Tabla 5.4: Frecuencias de los modos gravitacionales de la estrella KIC 11145123 según el modelo encontrado por Kurtz et al. (2014).

Frecuencia	2x2	3x3	3x4	4x4	Esf. 1	Esf. 2	L-S
1.209							
1.245							
1.284							
1.325							
1.369	✓	✓	✓	✓	✓	✓	✓
1.416	✓	✓	✓	✓	✓	✓	✓
1.467							
1.521		✓	✓				
1.579							
1.642							
1.712							
1.786							
1.868							
1.957							
2.057							

Una de las razones del pobre desempeño en la modelación de esta estrella puede ser la forma del espacio de características. Para espacios de diferentes dimensiones, asociados a diversos modelos de Neuropercolación el desempeño de PSO tanto en su valor mínimo, como media y mediana entrega valores similares. Esto puede sugerir que se presentan múltiples mínimos locales ubicados a través de todo el espacio, pero que los mínimos globales no se encuentran cercanos a estos mínimos locales, o se encuentran en zonas muy restringidas. Para evitar este problema se necesita mantener una mayor variabilidad en la población de PSO, por lo que se podrían utilizar técnicas de *annealing*, evolución separada de diversas poblaciones, u otras técnicas.

Otra razón posible para el desempeño encontrado es la cantidad acotada de información utilizada para la modelación. Dada la complejidad de la señal y sus múltiples componentes armónicas, el uso de únicamente 2000 datos en la curva de luz es una limitante para la modelación.

El mejor modelo, de estructura de grilla de 3x3 ayuda a dilucidar las razones por las que no se obtienen mejores modelos al realizar la optimización. Si se observa la figura 5.12 se puede concluir que las dos frecuencias principales asociadas a los modos gravitacionales (baja frecuencia) se encuentran representadas de forma clara, lo que es consistente con la información presentada en la tabla 5.4. Adicionalmente se observa que una de las componentes de alta frecuencia también es identificada. La otra componente se sabe que no fue modelada, dado que se encuentra presente en los residuos (ver figura 4.16).

Al enfocar la atención en las frecuencias de origen gravitacional, en la figura 5.13 se aprecian las tres frecuencias encontradas bajo el margen de 0.5 % de error relativo. Adicionalmente se observa una frecuencia adicional, cercana a la componente de 1.642 días^{-1} , pero que no se considera como identificada por tener un error mayor. Es fácil ver que la mayoría de estas frecuencias se pasan por alto completamente en el periodograma de Lomb-Scargle de la señal original. Es posible que este mismo defecto sea una de las razones por las cuales no se identifican más frecuencias en el modelo de Neuropercolación.

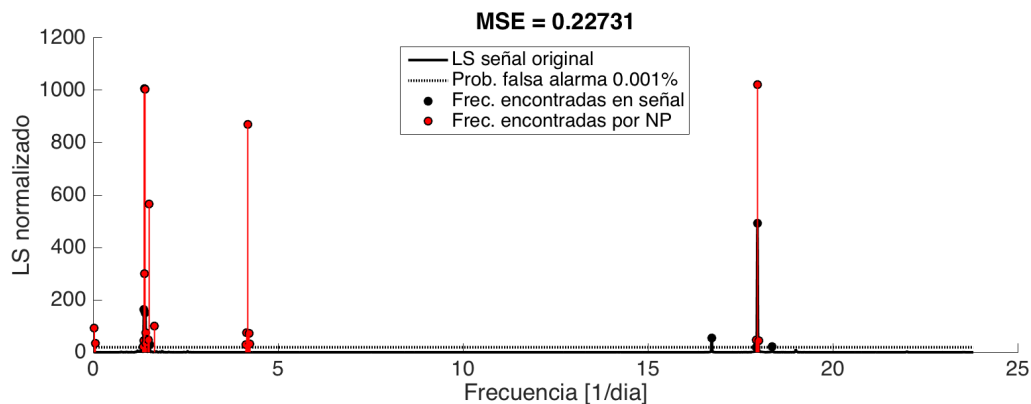


Figura 5.12: Espectro completo de frecuencias encontradas para lacurva de luz KIC 11145123 con el mejor modelo de 3x3.

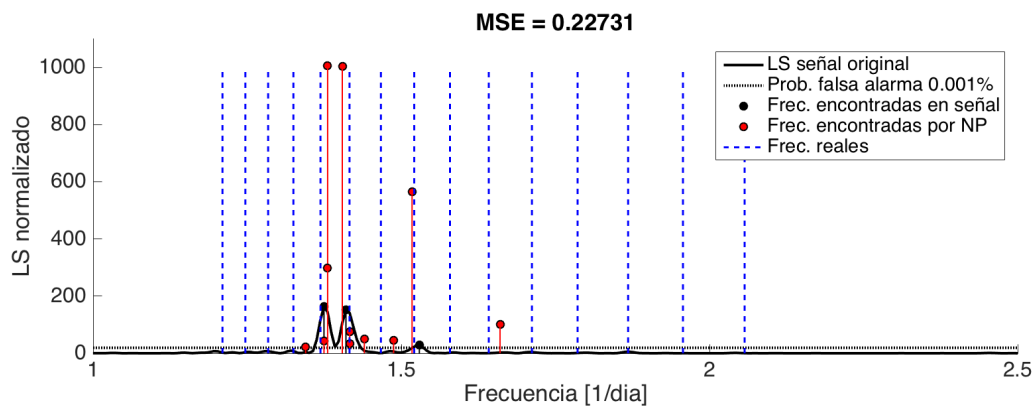


Figura 5.13: Frecuencias asociadas a modos gravitacionales encontradas para la curva de luz KIC 11145123 con el mejor modelo de 3x3.

5.3. Multiplicidad de los modelos

Es relevante señalar que aún para una señal generada mediante un modelo específico de Neuropercolación, un modelo distinto al generativo puede tener ajuste perfecto. Esto debido a que el orden de los osciladores no se encuentra fijo, y un modelo puede ser equivalente a otro, teniendo los mismos parámetros, pero los osciladores en posiciones diferentes. Adicionalmente, un modelo con duplicidad de osciladores, u osciladores aislados en puntos fijos con valor cero, pueden entregar los mismos resultados que un modelo sin tales redundancias. Este punto refleja tanto fortalezas como debilidades de la metodología de modelación. Por una parte permite una mayor flexibilidad a la hora de modelar, pero por otra no asegura resultados únicos ni óptimos globales.

Un ejemplo de lo señalado se observa en la figura 5.14, donde la topología de las subfiguras 5.14(a) y 5.14(b) es la misma, una grilla de 2x2, pero la posición de los osciladores es rotada. Esto puede llegar a considerarse como el mismo modelo, dependiendo de la forma con la que se trabajen los modelos. Si se consideran como individuos en una población de PSO, éstos se verían como individuos distintos con el mismo *fitness*. La subfigura 5.14(c) muestra los mismos osciladores, pero organizados en una grilla de 1x4 con condiciones toroidales.

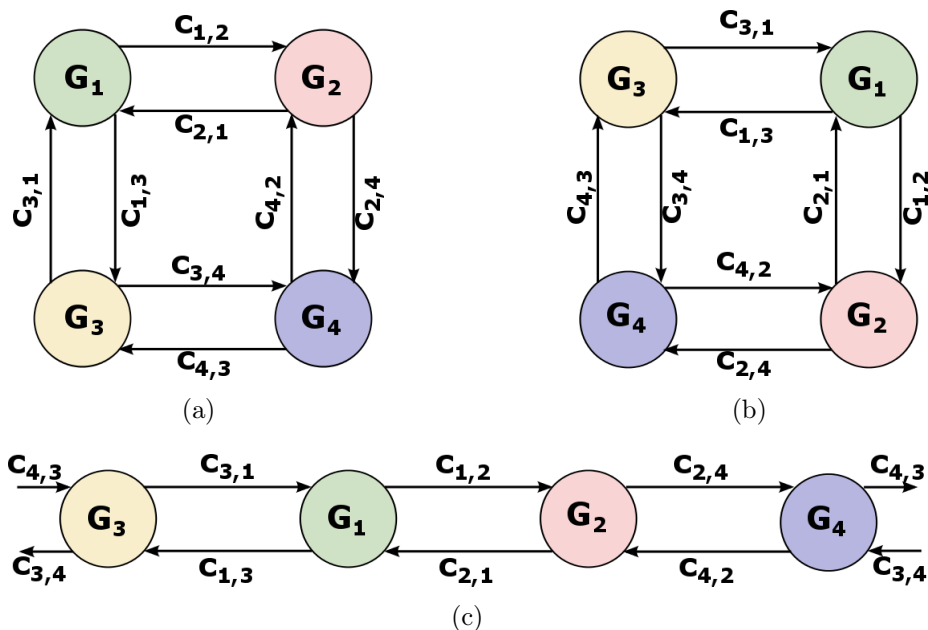


Figura 5.14: Ejemplo de multiplicidad de modelos de Neuropercolación.

5.4. Capacidad de predicción

Una de las posibilidades de exploración a futuro en estos modelos es la capacidad de realizar predicciones. A pesar de que el desarrollo realizado no tiene como objetivo principal la predicción, los resultados señalan que podría ser aplicado de esta manera. Para ello, vemos que aquellas señales con las que se obtienen estimaciones de muy bajo Error Cuadrático Medio, como por ejemplo la de Amplitud Modulada (ver figura 5.15), permiten una predicción también muy acertada. Pero señales más complejas como la simulación de curva de luz de estrella Gamma Doradus (ver figura 5.16) no logran un adecuado desempeño en la predicción.

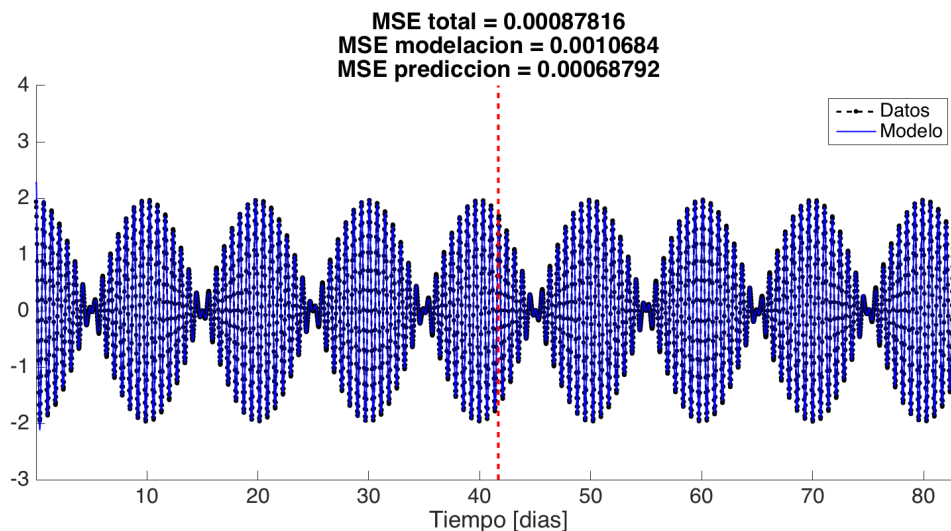


Figura 5.15: Predicción para señal de Amplitud Modulada. La línea roja señala el límite entre los datos utilizados para encontrar el modelo y la predicción.

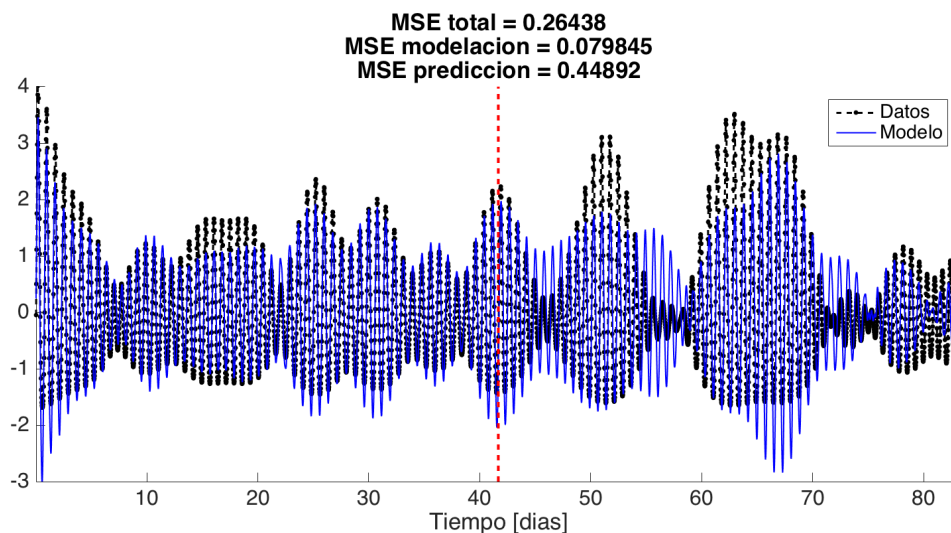


Figura 5.16: Predicción para curva de luz simulada de estrella Gamma Doradus utilizando el mejor modelo de grilla de Neuropercolación con tamaño 5x5. La línea roja señala el límite entre los datos utilizados para encontrar el modelo y la predicción.

Otro punto a considerar si se quiere profundizar en el tema de predicción es la cantidad de datos a utilizar. Como ya fue mencionado anteriormente, para obtener buenos modelos de señales complejas se requieren una mayor cantidad de información. Sin un modelo adecuado, la predicción no será acertada.

5.5. Transferencia de información e información *a priori*

El uso de información a priori, si bien es un paso lógico a considerar, no es tan simple de implementar en la práctica. El principal tipo de información que podría interesar en el contexto del problema planteado en este trabajo es aquella relacionada con la frecuencia de pulsación de las estrellas. Incluso algún tipo de estimación del rango de frecuencias de una estrella en particular podría resultar de utilidad. Ciertamente, información sobre tal aspecto existe para varias Gamma Doradus.

La dificultad con el uso de esta información es como aplicarla de manera efectiva al modelo. Se podrían fijar los períodos de algunos osciladores al iniciar la población de PSO, para lo cual habría que estudiar la distribución de frecuencias que puede presentar un oscilador en específico. Este estudio puede ser realizado a través de los conjuntos KII. Si se fija el valor del tamaño de la vecindad (d), entonces se pueden analizar las frecuencias que producen diferentes combinaciones de valores de los parámetros p y ω . Luego de esto surgen los problemas, cuyas soluciones no son triviales:

- **Conexiones entre osciladores:** A pesar de que un conjunto KII, al que llamamos un *oscilador* en el contexto de un modelo de Neuropercolación, tenga una cierta frecuencia cuando se considera como un modelo independiente, al incluirlo dentro de un modelo más grande las interconexiones con otros osciladores afectan su comportamiento. El efecto que tenga la conexión con otros osciladores puede variar desde un cambio de fase, un cambio de forma, o una alteración total de la frecuencia a la cual oscila. Tanto los parámetros de los osciladores a los cuales se encuentra conectado como el valor de la misma conexión tienen influencia en la conducta final.
- **Posición del oscilador:** Debido a las conexiones topológicas que presentan los osciladores, la posición que específica que tenga un oscilador en una grilla es parte fundamental del modelo. Aunque diversas topologías puedan ser análogas, esto solamente ocurre cuando todas las conexiones topológicas son iguales.

De resolverse los problemas mencionados, se podría utilizar información *a priori* y también realizar transferencia de información desde modelos pequeños de Neuropercolación hacia estructuras de mayor tamaño. Este proceso se basa en la idea de que si alguna parte de la estructura de una señal ya se encuentra modelada, entonces tal información puede ser útil para comenzar una búsqueda en un espacio de características más grande. Este procedimiento podría llegar a mejorar el desempeño de la búsqueda de PSO, evitando mínimos locales, pero también podría mejorar el tiempo computacional requerido, dado que modelos más pequeños presentan un menor costo computacional.

5.6. Costo Computacional

El mayor inconveniente asociado a la metodología propuesta corresponde al alto tiempo computacional que requiere. Este costo computacional se encuentra relacionado a más de un aspecto de la metodología, los cuales pueden ser abordados de manera de lograr un proceso más eficiente.

La primera razón de la complejidad computacional, y probablemente el punto más rígido en cuanto a posibilidades de optimización, corresponde a la simulación de cada modelo de Neuropercolación analizado durante la optimización. Estas simulaciones se realizan de forma iterativa, dado que la evolución de las probabilidades depende estrictamente de los estados de activación previos (ver ecuaciones en la sección 2.7.1). Las probabilidades de activación de cada oscilador deben ser calculadas y existen dependencias entre los estados de los osciladores debido a las interconexiones entre conjuntos KII. Por este motivo, sólo los estados de los múltiples osciladores para un tiempo determinado pueden ser calculados de manera independiente. En la implementación realizada el cómputo en cada instante de tiempo se hace en forma matricial, y dado el bajo número de osciladores con los que se trabaja (para este trabajo de tesis, $N < 30$) una implementación con computación paralela no aportaría mayores ventajas.

Si se analiza la implementación usada (ver anexo 7.2.3) se tienen tres ciclos *for* anidados: uno por cada instante de tiempo a simular, otro por la cantidad de osciladores, y otro asociado al tamaño de la vecindad. El último ciclo mencionado es el que calcula las probabilidades de activación, y al ser combinatorial determina la mayor cantidad de tiempo de cálculo.

La segunda razón de la complejidad computacional puede ser más sencilla de abordar, y corresponde al método de optimización. En el trabajo aquí presentado, la única herramienta de optimización utilizada fue Optimización por Enjambre de Partículas. Al utilizar este método con una población de partículas de N_{POP} y un número de iteraciones de N_{GEN} , entonces se deben realizar $N_{POP} \times N_{GEN}$ evaluaciones de la función de *fitness*, y por lo tanto la misma cantidad de simulaciones de modelos KIII. Una mejor inicialización o un método de optimización que alcance resultados similares para un número menor de evaluaciones de la función objetivo reduciría inmediatamente el tiempo computacional necesario para la modelación.

6 Conclusiones

La familia de modelos de Neuropercolación es una herramienta muy poderosa a la hora de representar señales complejas utilizando como base la teoría de autómatas celulares probabilísticos. En grillas de gran tamaño permiten dar una interpretación espacio-temporal a los fenómenos dinámicos que éstas representan. En el caso de representaciones de menor tamaño no se puede realizar tal interpretación, pero se mantiene la posibilidad de analizar las dinámicas existentes entre los diversos osciladores que comprenden el modelo.

En este trabajo fueron estimados los parámetros de diversos modelos de Neuropercolación con el objetivo de modelar series de tiempo, en particular aquellas relacionadas con la luminosidad observable de las estrellas, también llamadas curvas de luz. Algunas estrellas, entre ellas la familia de Gamma Doradus, son estrellas que varían su luminosidad en el tiempo. En particular este tipo de estrellas presentan variaciones periódicas y cuasi-periódicas, siendo el estudio de sus períodos un tema relevante para el área de la Asterosismología. Con esta motivación, se propuso el trabajo de modelar estas curvas de luz para descomponerlas en otras con espectros de frecuencia más simples, de manera de poder encontrar con precisión los períodos asociados al fenómeno de oscilación.

La estimación de parámetros de un modelo de Neuropercolación fue realizada utilizando Optimización por Enjambre de Partículas. Con un conjunto de parámetros dado es posible simular una serie de tiempo regular que puede ser comparada con una señal real. Al utilizar PSO para la estimación de parámetros, los resultados del modelo dependen de la inicialización de la población, siendo distintos en cada repetición del algoritmo.

Para estudiar la capacidad y robustez de la estimación de modelos de Neuropercolación se realizaron múltiples pruebas sobre una señal simple, teniendo conocimiento que su modelamiento era factible. La señal elegida fue de Amplitud Modulada, y se estudió el efecto sobre los resultados de realizar la estimación en presencia de datos perdidos y ruido blanco. Se concluye que a pesar de existir una pérdida en la robustez de la modelación, es factible obtener modelos representativos de la señal original, aún con un alto nivel de ruido o muchos datos perdidos. Sin embargo, aún cuando el conjunto de datos sea perfecto, puede ocurrir que una única estimación de parámetros entregue un modelo con un bajo nivel de ajuste, debido a presencia de diversos mínimos locales en el espacio de búsqueda. Debido a esto es necesario realizar múltiples simulaciones en búsqueda del mejor modelo. Los modelos obtenidos con menores grados de ajuste correspondientes a óptimos locales en el espacio de características logran representar una única frecuencia en vez de las dos buscadas para esta señal.

Dado que el tipo de señales reales que se quiere modelar tienen características específicas, siendo ricas en su espectro de frecuencia y con componentes periódicas de períodos cercanos, se simularon curvas de luz con estas características. La ventaja es que en estas simulaciones se conocen los períodos que se quieren encontrar, sirviendo como punto de referencia de la capacidad de esta familia de modelos. La modelación realizada en las curvas de luz de estrellas Gamma Doradus que fueron simuladas muestra que, dada una cantidad de osciladores similar a la cantidad de frecuencias que se quiere representar, gran parte de la base de frecuencias puede ser recuperada al analizar los osciladores por separado. Esta característica supera la capacidad de obtención de frecuencias del periodograma de Lomb-Scargle aplicado directamente a la señal. Por otro lado, se concluye que un modelo sobre-dimensionado, aunque pueda presentar un ajuste de MSE mejor, puede perder fiabilidad en la presencia de frecuencias relevantes en los osciladores.

Finalmente la metodología fue probada sobre señales reales de luminosidad de estrellas Gamma Doradus con datos obtenidos del proyecto Kepler, un telescopio espacial. Estos datos presentan un muestreo casi regular y con ruido muy bajo, características muy deseables para la modelación. Se prueban 3 curvas de luz distintas, de diversa complejidad. En el primer caso, de la estrella KIC 9240041, se lograron buenos niveles de ajuste, representando en gran parte la dinámica de la señal, aún con modelos pequeños. En el caso de la estrella KIC 10224094 no se obtuvieron resultados comparables al anterior caso, logrando ajustes con un MSE superior en un grado de magnitud. Esta estrella aparenta tener un espectro con presencia de múltiples frecuencias que no lograron ser representadas, pero que permanece como un desafío para trabajos futuros. Para el caso de esta estrella en particular, y para el caso de señales con múltiples componentes espectrales en general, es posible que el uso de una mayor cantidad de datos en la modelación facilite la búsqueda de mejores modelos. En el último caso, de la estrella KIC11145123 se obtienen resultados mezclados, donde a pesar de rescatar ciertos comportamientos relevantes de la estrella como la presencia de frecuencias altas y bajas, múltiples componentes no son identificadas. Dada la complejidad inherente en el espectro de esta señal, una potencial metodología con transferencia de información donde modelos pequeños ya entrenados se utilizan como punto de partida para la búsqueda de modelos de mayor tamaño, podría llegar a ser útil para mejorar los resultados obtenidos.

Por otro lado, dadas las pruebas realizadas con las diferentes topologías propuestas se pudo observar que los modelos esféricos propuestos pueden o no presentar mejores tasas de ajuste en comparación a modelos planos con similar cantidad de osciladores. En algunos casos se obtuvieron ajustes mínimos mejores, aunque en otros su desempeño era igual o peor. Los desempeños medios y medianos en múltiples repeticiones del algoritmo de optimización para modelos esféricos casi siempre resultaban peores que los modelos planos de tamaño similar.

De forma general, se concluye que los modelos de Neuropercolación presentan un gran poder de representación de señales complejas, con la característica añadida de la obtención de una descomposición. Por el momento siguen existiendo desafíos a la hora de utilizar estas estructuras para modelación directa de señales, aunque hay indicios de los beneficios que puede tener el realizar tal esfuerzo. Modificaciones en el proceso presentado, tanto en aspectos computacionales como en el proceso mismo de optimización podrían llevar a grandes mejoras en los resultados aquí discutidos.

Específicamente, mediante los resultados obtenidos en las pruebas planteadas en este trabajo, se concluye que la modelación de señales de amplitud modulada fue exitosa, logrando encontrar modelos robustos al ruido y escasez de datos. Para los casos analizados en este trabajo los modelos logran evitar exitosamente el sobre-ajuste al ruido. En el caso de las curvas de luz modeladas de Gamma Doradus se obtuvieron resultados prometedores. Si bien los ajustes de mínimos cuadrados no lograron niveles óptimos en la modelación, fue posible recuperar la base de frecuencias en algunos de los mejores modelos y construir un espaciamiento de periodos similar al teórico. Para las curvas de luz reales de estrellas Gamma Doradus los resultados fueron mezclados, logrando exitosamente modelar una de las curvas, mientras que las otras dos analizadas presentaron errores de mayor magnitud. En el caso de las curvas de luz con componentes de frecuencia conocidas en su espectro, no se logró reconstruir la base completa de frecuencias, sino únicamente rescatar las principales. Tanto el uso de una mayor cantidad de datos a la hora de modelar, como mejoras en el método planteado podrían disminuir estos errores.

6.1. Trabajo Futuro

Esta tesis presenta únicamente los primeros pasos de un estudio que pretende la obtención de mejores resultados, que faciliten el estudio e interpretación de señales de alta complejidad, en particular en el área astronómica.

De esta manera, los desafíos a trabajar en el futuro incluyen los siguientes,

- **Disminuir el tiempo de cómputo:** Como fue discutido, la complejidad computacional asociada a la simulación de un modelo Neuropercolación KIII es bastante alta, y aumenta tanto con la cantidad de osciladores seleccionados como con la cantidad de puntos a simular. Es importante dedicar trabajo al problema del tiempo de cómputo para que la aplicación de estos modelos sea factible en presencia de una gran cantidad de datos. El aumento de la velocidad de cálculo puede realizarse mediante una mejor implementación, por ejemplo utilizando paralelismo en tarjetas gráficas, o utilizando aproximaciones para disminuir la complejidad computacional del algoritmo de simulación.
- **Explorar otros métodos de optimización:** Debido a los objetivos de este trabajo, sólo Optimización por Enjambre de Partículas fue utilizado para adaptar los parámetros de los modelos de Neuropercolación para ajustar las simulaciones a los datos. Esto no significa que tal metodología sea la más adecuada para el problema planteado, y la exploración de otros métodos puede llevar a mejores resultados.
- **Utilización de una mayor cantidad de datos para el ajuste de los modelos:** Se observó que para señales con un gran número de componentes de frecuencia en su espectro, la cantidad de información utilizada fue insuficiente para recuperar las características de la señal. El uso de una mayor cantidad de datos para obtener los modelos, o la utilización de ensambles de modelos obtenidos con diferentes sub-conjuntos de datos es un trabajo que debe ser abordado.

- **Estudiar el sobre-ajuste y generalización de los modelos obtenidos:** Si se quiere que los modelos representen la dinámica de las señales, sin sobre-ajustarse a los datos con los que se obtienen estos modelos, es necesario realizar estudios más profundos sobre su capacidad de generalización.
- **Métodos de comparación:** La mayoría de las señales astronómicas, a diferencia de aquellas estudiadas en el presente trabajo, no se caracterizan por poseer un muestreo regular. Más aún, son altamente irregulares. Si se encontrara una manera de comparar estas señales altamente regulares con un modelo de Neuropercolación regular de manera efectiva, se podría facilitar el estudio de éstas.
- **Cómputo de espectros:** A pesar de que Lomb-Scargle sea un método utilizado ampliamente, no es necesariamente el mejor en identificar múltiples frecuencias. El analizar la metodología con otro algoritmo de cálculo de espectros sería útil para observar si las ventajas comparativas observadas en este trabajo se mantienen o no.
- **Transferencia de información:** Al inicializar cada búsqueda del método de optimización de manera aleatoria se ignora la información que se pueda conocer de antemano sobre el modelo que se quiere encontrar. Si se tiene un modelo de menor tamaño ya ajustado, se debería poder utilizar como punto de partida para acotar el espacio de búsqueda en un modelo de mayor envergadura. De esta manera se podría mejorar el rendimiento de búsqueda en los espacios de variables inherentemente más grandes de aquellos modelos con mayor cantidad de osciladores.

Bibliografía

- [1] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. Cambridge University Press, January 2004.
- [2] P. Glendinning, *Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations*, ser. Cambridge Texts in Applied Mathematics. Cambridge University Press, 1994.
- [3] A. M. Samoilenko, “Quasiperiodic oscillations,” vol. 2, no. 5, p. 1783, 2007.
- [4] W. Lewin, J. Van Paradijs, and M. Van der Klis, “A review of quasi-periodic oscillations in low-mass x-ray binaries,” *Space Science Reviews*, vol. 46, no. 3-4, pp. 273–377, 1988. [Online]. Available: <http://dx.doi.org/10.1007/BF00212242>
- [5] L. A. Balona, K. Krisciunas, and A. W. J. Cousins, “Gamma-Doradus - Evidence for a New Class of Pulsating Star,” *Monthly Notices of the Royal Astronomical Society*, vol. 270, p. 905, Oct. 1994.
- [6] A. B. Kaye, G. Handler, K. Krisciunas, E. Poretti, and F. M. Zerbi, “ γ Doradus Stars: Defining a New Class of Pulsating Variables,” *Publications of the Astronomical Society of the Pacific*, vol. 111, pp. 840–844, 1999.
- [7] M. W. Feast and A. R. Walker, “Cepheids as distance indicators,” *Annual review of astronomy and astrophysics*, vol. 25, pp. 345–375, 1987.
- [8] Simon, N. R. and Clement, C. M., “A provisional RR Lyrae distance scale,” *Astrophysical Journal*, vol. 410, pp. 526–533, Junio 1993.
- [9] C. Aerts, J. Christensen-Dalsgaard, and D. Kurtz, *Asteroseismology*. Springer, 2010.
- [10] M.-P. Bouabid, J. Montalbán, A. Miglio, M.-A. Dupret, A. Grigahcène, and A. Noels, “Theoretical seismic properties of pre-main sequence γ Doradus pulsators,” *Astronomy & Astrophysics*, vol. 531, p. A145, 2011.
- [11] N. R. Lomb, “Least-squares frequency analysis of unequally spaced data,” *Astrophysics and Space Science*, vol. 39, pp. 447–462, Feb. 1976.
- [12] J. D. Scargle, “Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data,” *Astrophysical Journal*, vol. 263, pp. 835–853,

Dec. 1982.

- [13] S. Wolfram, “Computation theory of cellular automata,” *Comm. Math. Phys.*, vol. 96, no. 1, pp. 15–57, 1984. [Online]. Available: <http://projecteuclid.org/euclid.cmp/1103941718>
- [14] R. Kozma, “Neuropercolation,” *Scholarpedia*, vol. 2, no. 8, p. 1360, 2007.
- [15] R. Kozma, M. Puljic, P. Balister, B. Bollobás, and W. J. Freeman, “Phase transitions in the neuropercolation model of neural populations with mixed local and non-local interactions,” *Biological Cybernetics*, vol. 92, 2005.
- [16] W. J. Freeman, *Mass Action in the Nervous System: Examination of the Neurophysiological Basis of Adaptive Behavior Through the EEG*. Academic Press, 1975.
- [17] R. Kozma and W. J. Freeman, *Cognitive Phase Transitions in the Cerebral Cortex - Enhancing the Neuron Doctrine by Modeling Neural Fields*. Springer, 2016.
- [18] J. Kennedy and R. Eberhart, “Particle swarm optimization,” *Proceedings of IEEE International Conference on Neural Networks*, pp. 1942–1948, 1995.
- [19] M. Dorigo, M. A. M. d. Oca, and A. Engelbrecht, “Particle swarm optimization,” vol. 3, no. 11, p. 1486, 2008.
- [20] T. Van Reeth, A. Tkachenko, C. Aerts, P. I. Pápics, P. Degroote, J. Debosscher, K. Zwintz, S. Bloemen, K. De Smedt, M. Hrudkova, G. Raskin, and H. Van Wincel, “Detecting non-uniform period spacings in the Kepler photometry of γ Doradus stars: methodology and case studies,” *Astronomy & Astrophysics*, vol. 574, p. A17, Feb. 2015.
- [21] NASA. (2015, September) Kepler: A search for habitable planets. [Online]. Available: <http://kepler.nasa.gov/>
- [22] Strasbourg astronomical data center. [Online]. Available: <http://cds.u-strasbg.fr/>
- [23] F. Takens, *Dynamical Systems and Turbulence, Warwick 1980: Proceedings of a Symposium Held at the University of Warwick 1979/80*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, ch. Detecting strange attractors in turbulence, pp. 366–381.
- [24] A. W. J. Cousins and P. R. Warren, “Variable Stars Observed During the Cape Bright Star Programme,” *Monthly Notes of the Astronomical Society of South Africa*, vol. 22, p. 65, 1963.
- [25] M.-A. Dupret, A. Grigahcène, R. Garrido, J. De Ridder, A. Moya, J.-C. Suárez, R. Scuflaire, M. Gabriel, and M.-J. Goupil, “Theoretical Aspects of g-mode Pulsations in gamma Doradus Stars,” *Memorie della Società Astronomica Italiana*, vol. 77, p. 366, 2006.
- [26] M.-P. Bouabid, M.-A. Dupret, S. Salmon, J. Montalbán, A. Miglio, and A. Noels, “Effects of the coriolis force on high-order g modes in γ doradus stars,” *Monthly Notices of the*

Royal Astronomical Society, vol. 429, pp. 2500 – 2514, 2013.

- [27] G. Handler, “Asteroseismology,” in *Planets, Stars and Stellar Systems*, T. Oswalt and M. Barstow, Eds. Springer Netherlands, 2013, pp. 207–241. [Online]. Available: http://dx.doi.org/10.1007/978-94-007-5615-1_4
- [28] M. Tassoul, “Asymptotic approximations for stellar nonradial pulsations,” *Astrophysical Journal Supplement Series*, vol. 43, pp. 469–490, Aug. 1980.
- [29] R. H. D. Townsend, “Surface trapping and leakage of low-frequency g modes in rotating early-type stars — ii. global analysis,” *Monthly Notices of the Royal Astronomical Society*, vol. 319, no. 1, pp. 289–304, 2000.
- [30] A. Miglio, J. Montalbán, A. Noels, and P. Eggenberger, “Probing the properties of convective cores through g modes: high-order g modes in SPB and γ Doradus stars,” *Monthly Notices of the Royal Astronomical Society*, vol. 386, pp. 1487–1502, May 2008.
- [31] J. Lafler and T. D. Kinman, “An RR Lyrae Star Survey with the Lick 20-INCH Astrograph II. The Calculation of RR Lyrae Periods by Electronic Computer.” *Astrophysical Journal Supplement*, vol. 11, p. 216, Junio 1965.
- [32] Lennart Lindegren and Dainis Dravins, “The fundamental definition of radial velocity,” *Astronomy & Astrophysics*, vol. 401, no. 3, pp. 1185 – 1201, Abril 2003.
- [33] M. Catelan and H. A. Smith, *Pulsating Stars*. Wiley-VCH, 2015.
- [34] Ferraz-Mello, S., “Estimation of Periods from Unequally Spaced Observations,” *The Astronomical Journal*, vol. 86, no. 4, pp. 619–624, 1981.
- [35] Dworetzky, M. M., “A period-finding method for sparse randomly spaced observations of ‘How long is a piece of string?’,” *Monthly Notices of the Royal Astronomical Society*, vol. 203, pp. 917–924, Junio 1983.
- [36] Schwarzenberg-Czerny, A., “On the advantage of using analysis of variance for period search,” *Monthly Notices of the Royal Astronomical Society*, vol. 241, pp. 153–165, Noviembre 1989.
- [37] Pablo Huijse and Pablo A. Estevez and Pavlos Protopapas and Pablo Zegers and José C. Principe, “An Information Theoretic Algorithm for Finding Periodicities in Stellar Light Curves,” *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5135 – 5145, Octubre 2012.
- [38] Pavlos Protopapas and Pablo Huijse and Pablo A. Estévez and Pablo Zegers and José C. Príncipe and Jean-Baptiste Marquette, “A Novel, Fully Automated Pipeline for Period Estimation in the EROS 2 Data Set,” *The Astrophysical Journal Supplement Series*, vol. 216, no. 2, p. 25, 2015.
- [39] P. Huijse, “Finding periodicities in astronomical light curves using information theoretic learning,” Ph.D. dissertation, Universidad de Chile, 2014.

- [40] L. H. Koopmans, *The Spectral Analysis of Time Series*, 2nd ed., ser. Probability and Mathematical Statistics, Z. W. Birnbaum and E. Lukacs, Eds. Academic Press, 1995, vol. 22.
- [41] Norbert Wiener, “Generalized harmonic analysis,” *Acta Mathematica*, vol. 5, no. 1, pp. 117–258, 1930.
- [42] Cooley, J. W. and Tukey, O. W., “An Algorithm for the Machine Calculation of Complex Fourier Series,” *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, April 1965.
- [43] L. Eyer and P. Bartholdi, “Variable stars: Which nyquist frequency?” *Astronomy & Astrophysics*, vol. 135, no. 1, pp. 1–3, Febrero 1999.
- [44] M. Zechmeister and M. Kürster, “The generalised Lomb-Scargle periodogram: A new formalism for the floating-mean and Keplerian periodograms,” *Astronomy & Astrophysics*, vol. 496, no. 2, pp. 577 – 584, Marzo 2009.
- [45] M. Gardner, “Mathematical games: The fantastic combinations of john conway’s new solitaire game "life",” *Scientific American*, vol. 223, no. 4, pp. 120–123, October 1970.
- [46] S. R. Broadbent and J. M. Hammersley, “Percolation processes. I. Crystals and Mazes,” *Proceedings of the Cambridge Philosophical Society*, vol. 53, pp. 629–641, Jul. 1957.
- [47] B. Bollobás and O. Riordan, *Percolation*. Cambridge University Press, 2006.
- [48] D. Stauffer and A. Aharony, *Introduction to percolation theory*, 2nd ed. Taylor & Francis, 1992.
- [49] S. Janson, T. Luczak, T. Turova, and T. Vallier, “Bootstrap percolation on the random graph $g_{N,p}$,” *The Annals of Applied Probability*, vol. 22, no. 5, pp. 1989 – 2047, 2012.
- [50] W. J. Freeman and H. Erwin, “Freeman K-set,” vol. 3, no. 2, p. 3238, 2008.
- [51] R. Kozma and M. Puljic, “Hierarchical random cellular neural networks for system-level brain-like signal processing,” *Neural Networks*, vol. 45, 2013.
- [52] I. Beliaev and R. Kozma, “Time series prediction using chaotic neural networks on the cats benchmark,” *Neurocomputing*, vol. 70, pp. 2426–2439, 2007.
- [53] W. J. Freeman, *How Brains Make Up Their Minds*. Columbia University Press, 2001.
- [54] R. Kozma and M. Puljic, “Patter-based computing via sequential phase transitions in hierarchical mean fiel neuropercolation,” *Theoretical Computer Science*, pp. 101–110, 2015.
- [55] M. Clerc and J. Kennedy, “The particle swarm - explosion, stability, and convergence in a multidimensional complex space,” *Evolutionary Computation, IEEE Transactions on*, vol. 6, no. 1, pp. 58–73, Feb 2002.
- [56] NASA. (2015, December) Kepler and K2 science center. [Online]. Available:

<http://keplerscience.arc.nasa.gov/>

- [57] S. B. Howell, C. Sobeck, M. Haas, M. Still, T. Barclay, F. Mullally, J. Troeltzsch, S. Aigrain, S. T. Bryson, D. Caldwell, W. J. Chaplin, W. D. Cochran, D. Huber, G. W. Marcy, A. Miglio, J. R. Najita, M. Smith, J. D. Twicken, and J. J. Fortney, “The K2 Mission: Characterization and Early Results,” *Publications of the Astronomical Society of the Pacific*, vol. 126, pp. 398–408, abril 2014.
- [58] Data Analysis Working Group (DAWG), “Kepler data release 24 notes: Q0–Q17,” April 2015. [Online]. Available: https://archive.stsci.edu/kepler/release_notes/release_notes24/KSCI-19064-002DRN24.pdf
- [59] T. M. Brown, D. W. Latham, M. E. Everett, and G. A. Esquerdo, “Kepler Input Catalog: Photometric Calibration and Stellar Classification,” *Astrophysical Journal*, vol. 142, p. 112, Oct. 2011.
- [60] Kepler public catalogs. [Online]. Available: <https://archive.stsci.edu/pub/kepler/catalogs/>
- [61] D. Fraquelli and S. E. Thompson, *Kepler Archive Manual (KDMC-10008-005)*, NASA, June 2014. [Online]. Available: https://archive.stsci.edu/kepler/manuals/archive_manual.pdf
- [62] L. A. Balona, J. A. Guzik, K. Uytterhoeven, J. C. Smith, P. Tenenbaum, and J. D. Twicken, “The kepler view of γ doradus stars,” *Monthly Notices of the Royal Astronomical Society*, vol. 415, no. 4, pp. 3531–3538, 2011.
- [63] T. Van Reeth, A. Tkachenko, C. Aerts, P. I. Pápics, S. A. Triana, K. Zwintz, P. Degroote, J. Debosscher, S. Bloemen, V. S. Schmid, K. D. Smedt, Y. Fremat, A. S. Fuentes, W. Homan, M. Hrudkova, R. Karjalainen, R. Lombaert, P. Nemeth, R. Østensen, G. V. D. Steene, J. Vos, G. Raskin, and H. V. Winckel, “Gravity-mode period spacings as a seismic diagnostic for a sample of γ doradus stars from kepler space photometry and high-resolution ground-based spectroscopy,” *The Astrophysical Journal Supplement Series*, vol. 218, no. 2, p. 27, 2015. [Online]. Available: <http://stacks.iop.org/0067-0049/218/i=2/a=27>
- [64] Balona, L. A. and Daszyńska-Daszkiewicz, J. and Pamyatnykh, A. A., “Pulsation frequency distribution in δ Scuti stars,” *Monthly Notices of the Royal Astronomical Society*, vol. 452, no. 3, pp. 3073–3084, 2015.
- [65] Donald W. Kurtz, Hideyuki Saio, Masao Takata, Hiromoto Shibahashi, Simon J. Murphy, and Takashi Sekii, “Astero seismic measurement of surface-to-core rotation in a main-sequence A star, KIC 11145123,” *Monthly Notices of the Royal Astronomical Society*, vol. 444, no. 1, pp. 102–116, Agosto 2014.
- [66] P. Balister, B. Bollobas, and R. Kozma, “Large deviations for mean field models of probabilistic cellular automata,” *Random Structures & Algorithms*, vol. 29, no. 3, pp. 399–415, 2006.

- [67] T. D. Sauer, "Attractor reconstruction," vol. 1, no. 10, p. 1727, 2006.
- [68] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5 - 6, pp. 237 - 329, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0370157306004066>
- [69] P. Huijse, P. A. Estévez, P. Zegers, J. C. Príncipe, and P. Protopapas, "Period estimation in astronomical time series using slotted correntropy," *Signal Processing Letters, IEEE*, vol. 18, pp. 371 - 374, 2011. [Online]. Available: <http://arxiv.org/abs/1112.2962>
- [70] J. L. Christiansen, J. M. Jenkins, D. A. Caldwell, T. Barclay, S. T. Bryson, C. J. Burke, J. Campbell, J. Catanzarite, B. D. Clarke, J. L. Coughlin, F. Girouard, M. R. Haas, K. Ibrahim, T. C. Klaus, J. J. Kolodziejczak, J. Li, S. D. McCauliff, R. L. Morris, F. Mullally, E. V. Quintana, J. Rowe, A. Sabale, S. Seader, J. C. Smith, M. D. Still, P. G. Tenenbaum, S. E. Thompson, J. D. Twicken, and A. K. Uddin, *Kepler Data Characteristics Handbook (KSCI - 19040 - 004)*, NASA, May 2014. [Online]. Available: https://archive.stsci.edu/kepler/manuals/Data_Characteristics.pdf
- [71] M. N. Fanelli, J. M. Jenkins, S. T. Bryson, E. V. Quintana, J. D. Twicken, H. W. Wu, P. Tenenbaum, C. L. Allen, D. A. Caldwell, H. Chandrasekaran, J. L. Christiansen, B. D. Clarke, M. T. Cote, J. L. Dotson, R. Gilliland, F. Girouard, J. P. Gunter, J. Hall, M. R. Haas, K. Ibrahim, K. Kinemuchi, T. Klaus, J. Kolodziejczak, J. Li, P. Machalek, S. D. McCauliff, C. K. Middour, R. Morris, F. Mullally, S. Seader, J. C. Smith, M. Still, S. E. Thompson, A. K. Uddin, J. Van Cleve, and B. Wohler, *Kepler Data Processing Handbook (KSCI-19081-001)*, NASA, April 2011. [Online]. Available: https://archive.stsci.edu/kepler/manuals/KSCI-19081-001_Data_Processing_Handbook.pdf
- [72] B. Boashash, "Chapter 1 - time-frequency concepts," in *Time Frequency Analysis*, B. Boashash, Ed. Oxford: Elsevier Science, 2003, pp. 3 - 27. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/B97800804433550022X>
- [73] J. A. Guzik, A. B. Kaye, P. A. Bradley, A. N. Cox, and C. Neuforge, "Driving Gravity-Mode Pulsations in γ Doradus Variables," in *Delta Scuti and Related Stars*, ser. Astronomical Society of the Pacific Conference Series, M. Breger and M. Montgomery, Eds., vol. 210, 2000, p. 423.

7 Anexos

7.1. Ecuaciones de la variabilidad estelar

Estos desarrollos son los que se encuentran en *Asteroseismology* de Aerts, Christensen-Dalsgaard y Kurtz [9].

Para una estrella esférica y simétrica las soluciones de las ecuaciones de movimiento tienen desplazamientos en las direcciones (r, θ, ϕ) que se encuentran dadas por:

$$\xi_r(r, \theta, \phi, t) = a(r)Y_l^m(\theta, \phi) \exp(-i2\pi\nu t) \quad (7.1)$$

$$\xi_\theta(r, \theta, \phi, t) = b(r)\frac{\partial Y_l^m(\theta, \phi)}{\partial \theta} \exp(-i2\pi\nu t) \quad (7.2)$$

$$\xi_\phi(r, \theta, \phi, t) = \frac{b(r)}{\sin \theta} \frac{\partial Y_l^m(\theta, \phi)}{\partial \phi} \exp(-i2\pi\nu t) \quad (7.3)$$

$$(7.4)$$

En este conjunto de ecuaciones, ξ_r , ξ_θ y ξ_ϕ corresponden a los desplazamientos en las tres direcciones ortogonales, $a(r)$ y $b(r)$ son amplitudes y ν es la frecuencia de oscilación. $Y_l^m(\theta, \phi)$ son los armónicos esféricos, cuya expresión es la siguiente,

$$Y_l^m(\theta, \phi) = (-1)^m \sqrt{\frac{2l+1}{\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) \exp(im\phi) \quad (7.5)$$

$$P_l^m(\cos \theta) = \frac{1}{2^l l!} (1 - \cos^2 \theta)^{m/2} \frac{d^{l+m}}{d \cos^{l+m} \theta} (\cos^2 \theta - 1)^l \quad (7.6)$$

Los valores de $P_l^m(\cos \theta)$ corresponden a los Polinomios de Legendre, y se encuentran definidos en la ecuación 7.6.

7.2. Códigos en MATLAB

7.2.1. Código para simulación de modelo KI

```
function [a]=KI(T,w,d)
%Freeman KI set with mean field approximation
%w = probability of following majority
%T = ammount of points to simulate
%d = size of the neighborhood

D = floor(d/2);
a = zeros(1,T);

for t = 2:T

    Nw = 0; Ne = 0;
    aux = 0;

    for k = 0:D
        nck = nchoosek(d,k);
        Nw = Nw + nck*(a(t-1)^(d-k)*(1-a(t-1))^k - ...
            ... (1-a(t-1))^(d-k)*a(t-1)^k);
        Ne = Ne + nck*((1-a(t-1))^(d-k)*a(t-1)^k);
        aux = aux + nck*(d-2*k);
    end

    a(t)=w*Nw+Ne;

end

% Critical probability
w0 = 0.5+(2^(d-2))/(aux);
disp(['w0= ', num2str(w0)])
```

7.2.2. Código para simulación de modelo KII

```
function [a0 , a1]=KII(T,d,p,w)
% One oscillator (Freeman KII set)
% T = ammount of points to simulate
% d = size f the neighborhood
% p = inhibition strength
% w = probability of following the majority

D = floor(d/2);
a0 = zeros(1,T);
a1 = zeros(1,T);

for t = 2:T
    A0 = a0(t-1)*(1-p)+p*(1-a1(t-1));
    A1 = a1(t-1)*(1-p)+p*a0(t-1);

    Nw0 = 0; Ne0 = 0; Nw1 = 0; Ne1 = 0;
    aux = 0;

    for k = 0:D
        nck = nchoosek(d,k);
        Nw0 = Nw0 + nck*(A0^(d-k)*(1-A0)^k-(1-A0)^(d-k)*A0^k);
        Nw1 = Nw1 + nck*(A1^(d-k)*(1-A1)^k-(1-A1)^(d-k)*A1^k);

        Ne0 = Ne0 + nck*((1-A0)^(d-k)*A0^k);
        Ne1 = Ne1 + nck*((1-A1)^(d-k)*A1^k);

        aux = aux + nck*(d-2*k);
    end

    a0(t)=w*Nw0+Ne0;
    a1(t)=w*Nw1+Ne1;
end

% Critical probability
phi = (1-2*p)/(1-3*p+4*p^2-2*p^3);
w0 = 0.5+(2^(d-2))*phi/(aux);
disp([ 'w0= ', num2str(w0)])
```

7.2.3. Código para simulación de modelo KIII

```
function [a0, a1, connections] = KIII(T,d,p,w,c, percParams)
%Freeman KIII set with mean fields approximation
%Use: [a0, a1, connections] = KIII(T,d,p,w,c, percParams)
%
%percParams is a structure with the following attributes:
%percParams.N = Number of oscillators
%percParams.K = Connections per oscillator
%percParams.connections = ID of the connections
%percParams.comb = combinations [ nchoosek(percParams.d,i) ]. Improves
%
%           computational speed
%percParams.d = size of the neighborhood
%
%Parameters given by the user:
%T = steps for simulation
%d = vector with d values (Nx1)
%p = inhibition strength (vector length N)
%w = probability of following majority (vector length N)
%c = connection strengths (matrix NxK)

connections = percParams.connections;
N = percParams.N;
K = percParams.K;
comb = percParams.comb;

if isempty(connections)
    connections = zeros(N,K);

    for i = 1:N
        options = [1:i-1,i+1:N];
        options = options(randperm(length(options))));
        connections(i,:) = options(1:K);
    end
end

a0 = zeros(N,T);
a1 = zeros(N,T);

%Check if sizes are correct
if ~(isvector(p) && isvector(w)&& length(p)==length(w))
    disp('Size_of_vector_p_and_w_should_be_the_same:_Vector_of_length_N')
elseif ~isequal(size(c), [N,K])
    disp('Size_of_matrix_c_is_incorrect._Should_be_NxK')
elseif K>=N
    disp('N_must_be_greater_than_K')
```

```

end

for t = 2:T
    for j = 1:N
        A0 = a0(j,t-1)*(1-p(j)-sum(c(j,:))) + p(j)*(1-a1(j,t-1)) + ...
            ... (c(j,:)*a0(connections(j,:),t-1));
        A1 = (1-p(j))*a1(j,t-1)+p(j)*a0(j,t-1);

        Nw0=0; Ne0=0; Nw1=0; Ne1=0;
        D = floor(percParams.d/2);

        for k=0:1:D;
            nck = comb(k+1);
            %nck=nchoosek(d(j),k);

            tmp0=A0^(d(j)-k)*(1-A0)^k-(1-A0)^(d(j)-k)*A0^k;
            tmp1=A1^(d(j)-k)*(1-A1)^k-(1-A1)^(d(j)-k)*A1^k;
            mem0=(1-A0)^(d(j)-k)*A0^k;
            mem1=(1-A1)^(d(j)-k)*A1^k;

            Nw0=Nw0+nck*tmp0; Ne0=Ne0+nck*mem0;
            Nw1=Nw1+nck*tmp1; Ne1=Ne1+nck*mem1;
        end

        if rem(d(j),2)==0
            %nck = nchoosek(d(j), d(j)/2);
            nck = comb(d(j)/2+1);
            Nd0 = 0.5*nck*(A0*(1-A0))^(d(j)/2);
            Nd1 = 0.5*nck*(A1*(1-A1))^(d(j)/2);
        else
            Nd0 = 0;
            Nd1 = 0;
        end

        a0(j,t)=w(j)*Nw0+Ne0+Nd0;
        a1(j,t)=w(j)*Nw1+Ne1+Nd1;

    end
end
end

```

7.3. Identificación de frecuencias en modelos de Neuropercolación

Para obtener las frecuencias asociadas a cada oscilador, se analiza cada uno de éstos por separado utilizando el periodograma de Lomb-Scargle. Luego de tener cada periodograma independiente, se obtienen los máximos locales del periodograma, mientras cumplan con el umbral de falsa alarma establecido para las pruebas. Este proceso se ilustra en la figura 7.1, donde se muestran los cuatro osciladores de un modelo ajustado a una señal de Amplitud Modulada. En la figura, los gráficos de color negro representan las señales generadas por cada oscilador; los gráficos en azul corresponden a los periodogramas de Lomb-Scargle de cada oscilador; y en rojo se observan los máximos locales.

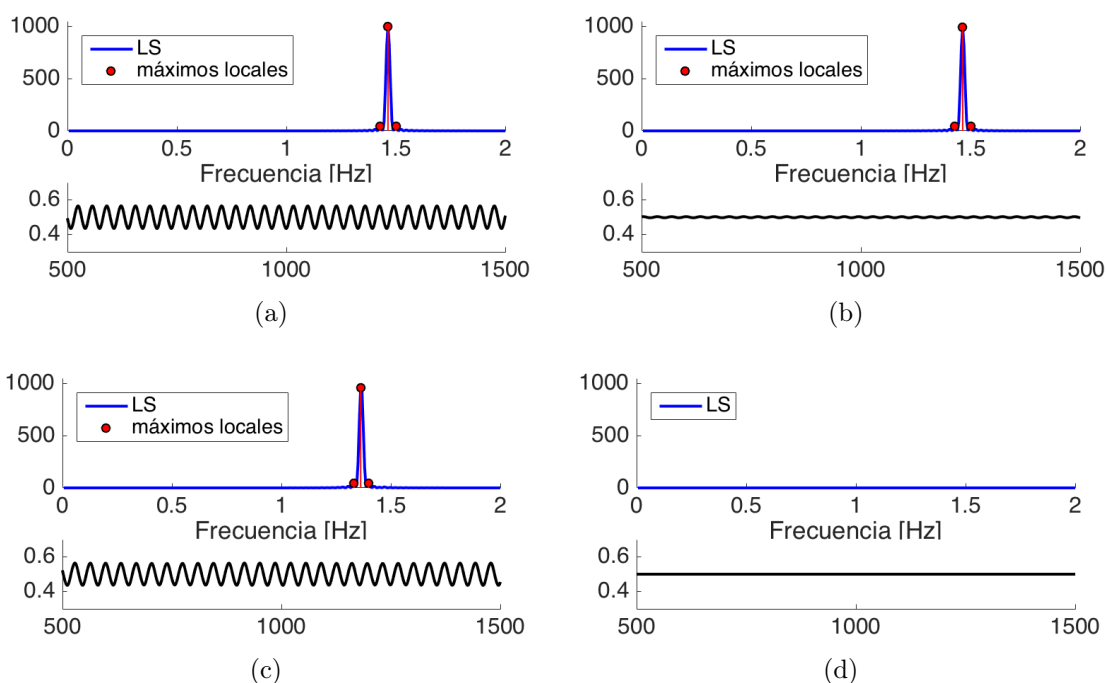


Figura 7.1: Extracción de frecuencias en modelos de Neuropercolación, mediante el análisis independiente de cada oscilador de los modelos. Las figuras 7.1(a), 7.1(b), 7.1(c) y 7.1(d) muestran los periodogramas de Lomb-Scargle sobre los cuatro osciladores de un modelo de Neuropercolación de estructura 2x2, para una señal simulada de Amplitud Modulada.