



Inconsistencias en el análisis armónico de series temporales de estrellas pulsantes observadas desde satélites

Javier Pascual Granado

Departamento de Física Estelar
Instituto de Astrofísica de Andalucía - CSIC

Tesis dirigida por
Rafael Garrido Haba
Juan Carlos Suárez Yanes

Programa Oficial de Posgrado en Física

Presentada en el departamento de
Física Teórica y del Cosmos de la UGR
para optar al grado de

Doctor en Astrofísica

Marzo 2014

Editor: Editorial de la Universidad de Granada
Autor: Javier Pascual Granado
D.L.: GR 1937-2014
ISBN: 978-84-9083-108-3

Al recuerdo,
del Abi y la Yaya

Harto más merecedores de predilección para el maestro
avisado serán aquellos discípulos un tanto
indómitos, desdeñosos de los primeros lugares,
insensibles al estímulo de la vanidad que,
dotados de rica e inquieta fantasía, gastan el
sobrante de su actividad en la literatura,
el dibujo, la filosofía y todos los deportes del
espíritu y del cuerpo. Para quien los sigue
de lejos, parece como que se dispersan y se disipan,
cuando, en realidad, se encauzan y fortalecen.
Corazones generosos, poetas a ratos, románticos
siempre... Al revés de los otros, al abandonar
las aulas es cuando realmente comienzan a estudiar.

-Santiago Ramón y Cajal —

“I don't believe in fate. You make your own opportunities.“

-Don Draper

Agradecimientos

Corría el año 86 y yo tenía 7 años. Vivíamos entonces en Sevilla y todavía se podían encontrar en esa época llanos desde donde observar el cielo abierto que no hubieran sido ocupados por la devastadora marea urbanizadora. Teníamos un pequeño telescopio refractor de 50 aumentos que mi padre compró para iniciarnos en la astronomía y salimos con él a observar el cometa Halley en uno de esos llanos enfrente de nuestro barrio. Era un telescopio de iniciación, sin embargo, con uno de características similares Galileo descubrió las lunas de Júpiter. Yo no llegaría a tanto pero para mí ese telescopio y el cometa Halley supusieron un descubrimiento que marcaría mi vida, pues a través de ellos y de mi padre conocí la astronomía. El fenómeno de la aparición del Halley en sí no me maravillaría tanto como la singularidad del evento, el estar viviendo algo que nadie tendría la oportunidad de vivir hasta dentro de 75 años. Movidado por ese interés, más tarde, indagaría en la literatura para saber más sobre el tema. Siempre he sido más de buscar las respuestas en los libros que en las personas, quizás por introversión. De todos modos, tenía la excusa perfecta ya que en casa mi padre tenía uno de los mejores libros de astronomía que se han escrito, *Astronomía Popular* de Camille Flammarion, en una edición bellamente ilustrada y corregida nada menos que por José Comas Solá.

A la fascinación por las noches estrelladas añadiría después un interés más intelectual cuando vi el documental *Cosmos* de Carl Sagan con mi padre. Creo que fue más o menos entonces cuando decidí que quería ser astrónomo, y aunque la adolescencia y la pos-adolescencia sembraron el desorden en mi vida, siempre tuve presente cuál había sido el principio y cuál debía ser el fin de mi dedicación. Años más tarde volvería a recordar ese principio cuando mi padre trajo a casa *Cometas*, el libro que Carl Sagan junto a su mujer, Ann Druyan, escribieron justo antes de la llegada del cometa Halley y que dedicaba una parte muy importante a este.

No quiero hacer un relato autobiográfico aquí pero quiero destacar que, a pesar de que

ellos nunca me hayan alentado a que hiciera esto, bastaba con ser un poco soñador para que con los estímulos de mi padre y el refuerzo moral y sentimental de mi madre me sintiera con deseos de explorar las estrellas. Si no fuera por ellos nunca hubiera acabado defendiendo esta tesis. Por eso, a ellos les dedico este trabajo. Gracias de todo corazón.

Quiero ahora dar las gracias especialmente a Rafael Garrido, mi director de tesis, quien ha sido mi mentor y ejemplo profesional durante todo este tiempo y con quien además de Ciencia en mayúsculas, he aprendido mucho de la sociedad científica. Ha sido un placer estar bajo su supervisión y bajo la del co-director, Juan Carlos Suárez, y compartir con ellos esta gran aventura. Gracias por la serenidad y temple que habéis mostrado siempre conmigo.

A mis hermanos Jordi y José Ramón, que me han dado su apoyo y que han compartido toda la caminata sobre brasas que es la tesis y han sido cómplices por haber tenido que vencer ellos sus propios desafíos a la vez que yo el mio.

Agradezco a Ahmed Grigahcène, quien se ha preocupado más por mi que yo mismo a veces, y es además una de las personas que más me han aportado y a quien más debo agradecer profesionalmente.

A Antonio García, que ha sido como un hermano mayor, y a Alba Fernández, que ha sido como una hermana menor (precoz, eso sí). A ambos no puedo agradecerles lo suficiente la ayuda que me han prestado y su amistad.

A Darío Díaz, pues un día, tomando un café conmigo cuando debíamos estar en una clase del Máster, me habló de un trabajo predoctoral sobre estrellas pulsantes en el departamento de Física Estelar del IAA, y con ello dio comienzo todo.

A todos los personajes del IAA como mi compañero de batalla el camarada Cortés, mi protectora Baba Marta, el infatigable Milan, el afable Víctor Muñoz, Ale, Raúl, Víctor Terrón, Juanma, Pablo Santos, y muchos otros (que están y que estuvieron) con quienes he compartido buenos momentos y espero poder compartir muchos más.

Y entre personajes, a Miguel Pío en especial, pues con él he compartido casi todo lo que se puede compartir y aún así sigue siendo mi amigo, lo que prueba que es un gran tipo.

Quiero agradecer también a todas las compañeras y compañeros del Grupo de Variabilidad Estelar, excelentes personas todas ellas y en especial a Cris, por compartir conmigo su excelente sentido musical, entre otras cosas, y a Juan Gutiérrez, que tantas veces se ha ofrecido a ayudarme y me ha dado consejo.

También quisiera extender el agradecimiento al personal del IAA que me ha acogido de manera tan cálida durante mi paso por el centro, y particularmente a Matilde Fernández. Así mismo no quiero dejar de mencionar aquí la profesionalidad y el buen hacer, a mi entendimiento, de los trabajadores del IAA, sobretodo el maravilloso equipo de divulgación llevado por Emilio y Silbia. Compartir centro con ellos ha sido para mi uno de los mayores motivos de orgullo durante estos años.

No quiero olvidarme de mis buenos colegas fuera del IAA. Por como han influido en mi modo de ver las cosas o por la decrepitud que han aportado a mi vida, no me olvido tampoco de Rau, Keni, Carol, Chus, Damián, Cécile, Romu, Vicky, Noe, Lore, y mi más antiguo compañero en el mundo de la astronomía, Jose el Maes.

¡Un brindis por todos vosotros!

Dedico estas últimas líneas a quien, en sus propias palabras, ha sufrido el mayor daño colateral de esta tesis, moja najzgodnija na svijetu Nikolina. Gracias porque a tu lado casi todos los días que ha durado este colosal esfuerzo han sido días felices. Y gracias también por hacer que me convierta al holmesianismo, porque el análisis de series temporales es mucho más emocionante cuando se coge la lupa y se estudian las huellas con astucia.

Índice

I	Introducción	5
1	Plan de Tesis	7
2	Marco general	11
2.1	Pulsación Estelar	11
2.1.1	Introducción a la astrosismología	11
2.1.2	Clasificación de estrellas pulsantes	16
2.1.3	astrosismología desde satélites	17
2.2	Técnicas de análisis de series temporales en astrosismología	19
2.3	Criterio de corrección científica	22
II	Eliminación de frecuencias espurias mediante reconstrucción de la señal	27
3	Efectos del muestreo en la detección de señal: la ventana espectral	31
4	Técnicas de relleno de huecos	41
4.1	Métodos de Interpolación	41
4.1.1	Interpolación lineal	42
4.1.2	Inpainting	43
4.1.3	El método de Fahlman y Ulrych (MFU)	44
4.1.4	Otros métodos	45
4.2	Modelos autoregresivos	47
4.3	El Método de Interpolación ARMA	50
4.3.1	Criterio para la elección del orden	50
4.3.2	Determinación de los coeficientes del modelo	53
4.3.3	Procedimiento de relleno	54
4.3.4	Flujo de datos	54

5	Resultados: series sintéticas y curvas de luz de CoRoT	57
5.1	Tests del método mediante simulaciones numéricas	57
5.1.1	Función periódica sinusoidal	58
5.1.1.1	Bajas frecuencias	58
5.1.1.2	Altas frecuencias	63
5.1.1.3	Modelo realista	71
5.2	Curvas de luz de CoRoT	81
5.2.1	Estrella tipo-solar: HD 49933	81
5.2.2	δ Scuti: HD 174966	83
5.2.3	Estrella Be: HD 51193	89
III	Estudio de la validez del desarrollo en serie de Fourier de una serie temporal	93
6	Consideraciones sobre la consistencia del análisis de frecuencias	97
6.1	La distribución gaussiana	98
6.2	Desviaciones sobre la gaussiana	99
6.3	Condiciones para el análisis armónico	102
7	El método de las conectividades	103
7.1	Conectividad	103
7.1.1	Concepto	103
7.1.2	Definición	105
7.1.3	El coeficiente de no-diferenciabilidad de una función	106
7.2	Métodos	107
7.2.1	Splines	107
7.2.2	ARMA	108
8	Analicidad de las series observadas	111
8.1	HD 174936	112
8.1.1	Modelo analítico para HD 174936	112
8.1.2	Análisis de las conectividades	114
8.1.3	SSE	116
8.2	KIC 006187665	118
8.2.1	SSE	119
8.2.2	Conectividades de orden 2	120
8.2.3	Análisis espectral de las conectividades	120
8.3	SoHO / GOLF	123

8.3.1	SSE	123
8.3.2	Análisis espectral de las conectividades	124
8.4	HD 181072	125
8.4.1	SSE	126
8.4.2	Análisis espectral de las conectividades	126
8.5	Discusión	127
IV Conclusiones y trabajo futuro		131
9 Conclusiones		133
10 Trabajo futuro		137
Apéndices		139
A El programa de análisis de frecuencias FRECS		141
Índice de figuras		149
Índice de tablas		157
Referencias		159

Resumen

Desde el descubrimiento en los años 60 de que la gran mayoría de las estrellas se hallan en un estado de inestabilidad pulsante, se vienen estudiando las pulsaciones como mecanismo para obtener información sobre la física y estructura de interiores estelares. El avance observacional logrado durante la última década gracias a misiones espaciales como CoRoT, MOST y *Kepler* nos ha proporcionado una ingente cantidad de datos sobre las estrellas observadas que ha permitido mejorar los modelos teóricos de estas. Al mismo tiempo, esos datos, analizados bajo las herramientas matemáticas clásicas y habitualmente utilizadas en el estudio de oscilaciones en todas las ciencias, están dando resultados que en un buen número de casos se hallan en disconformidad con las expectativas teóricas según el nivel de precisión alcanzado en las observaciones.

El procedimiento habitual en el análisis de las curvas de luz estelares se basa en una manipulación de los datos que introduce hipótesis de base como la descripción de series temporales en una base de funciones ortogonales, concretamente la descomposición en una base de Fourier.

Asimismo, para el tratamiento de datos procedentes de observaciones interrumpidas, se considera suficiente el uso de una interpolación lineal en los huecos cuando estos son de corta duración, o el uso de herramientas de análisis que asuman la no uniformidad del muestreo cuando los huecos son de mayor duración.

En cualquier caso, toda la interpretación física posterior depende de la coherencia interna del conjunto de hipótesis necesarias para su interpretación.

En este trabajo hemos abordado el análisis de los datos de la misión CoRoT desde un punto de vista no estándar. Para ello hemos desarrollado herramientas y técnicas de análisis que permitan llevar a cabo un estudio completamente neutral, en el sentido de que para poder validar hipótesis basadas en las observaciones, los datos deben estudiarse sin introducir sesgos.

Así pues, con la premisa de que el tratamiento de los datos es correcto siempre que no se altere la señal, se ha planteado una alternativa al uso de funciones analíticas como kernel para la reconstrucción de la señal en huecos, ya que esta técnica, usada

habitualmente, se basa en hipótesis ad hoc. En su lugar, se propone una técnica cuya predicción de valores futuros de la serie depende únicamente de la información contenida en la serie. Esta técnica basada en modelos ARMA se prueba en series sintéticas y sobre las curvas de luz de las estrellas observadas por el satélite CoRoT.

Una vez las curvas de luz han sido preprocesadas y se han corregido efectos sistemáticos y huecos, el paso habitual en el análisis es pasar la serie temporal al espacio de frecuencias mediante técnicas basadas en la transformada de Fourier para lograr la identificación de los modos de pulsación. La interpretación de los resultados, ya sea como frecuencias de pulsación, ruido u otro tipo de señal, varía según el caso y las hipótesis de base de las herramientas de análisis utilizadas. Dependiendo de cuáles sean estas se obtendrán distintas frecuencias significativas. Por tanto, asumiendo que los resultados de cualquier análisis científico han de ser objetivos y reproducibles, se ha desarrollado aquí una herramienta para la detección de frecuencias que se basa en el principio natural de que todo aquello distinguible de un proceso estocástico no correlado e independiente puede considerarse significativo. La eficiencia computacional de esta herramienta permite su uso masivo para analizar la cantidad de datos observacionales que se obtienen tanto de CoRoT como de Kepler.

El resultado principal de esta tesis es la demostración de que las funciones subyacentes a las series temporales de estrellas pulsantes observadas no son analíticas. La no analiticidad implica que las condiciones para que estas funciones sean desarrollables en series de Fourier no están garantizadas. En consecuencia, la necesaria coherencia interna del conjunto de hipótesis habitualmente admitidas como ciertas en el análisis de series temporales, no se cumple para el caso de las observaciones aquí analizadas. En conclusión, se demuestra que el análisis de Fourier de las series temporales de estrellas pulsantes observadas desde satélites presenta inconsistencias internas.

Parte I

Introducción



Capítulo 1

Plan de Tesis

“Omnia praeclara rara”...

Cicerón

La realización de esta tesis doctoral tuvo su comienzo hacia octubre de 2009 tras la finalización del Trabajo de Investigación Tutelada (TIT) del Máster de Métodos y Técnicas Avanzadas en Física (MTAF) de la Universidad de Granada (UGR). El satélite CoRoT, que en ese momento se dirigía hacia el final de su tercer año de vida con una cuota importante de hallazgos científicos en el campo de la astrosismología y los exoplanetas, abría asimismo la puerta a nuevas preguntas originadas en la interpretación de las observaciones y su contrastación con modelos teóricos sobre la estructura y evolución de las estrellas, como es un ejemplo el origen del abundante espectro de frecuencias encontrado en algunas estrellas pulsantes de masa intermedia que no se puede ajustar con los modelos teóricos conocidos de estas estrellas. Frente a la nueva problemática surgida se hace imperante la necesidad de reevaluar los métodos y la práctica del análisis de datos observacionales y de hallar una manera más rigurosa de comprobar las hipótesis aplicadas a las estrellas pulsantes. La labor realizada en la presente tesis busca la comprensión de la problemática surgida dentro del paradigma de la astrosismología a partir del desarrollo de nuevas técnicas y el análisis crítico de las técnicas habituales.

El problema fundamental de la astrosismología, que es la identificación de los modos de pulsación de las estrellas, cuya solución es necesaria para la inversión de los datos de la estructura estelar interna, requiere como paso inicial la detección de las frecuencias de pulsación. La detección lleva implícita la discriminación entre las frecuencias que provienen de pulsación estelar y las que no. Este es un problema en absoluto trivial que requiere una comprensión de los fenómenos implicados, un tratamiento correcto de los datos, y un examen profundo para verificar que se cumplen

1. PLAN DE TESIS

las condiciones tanto matemáticas como experimentales, requeridas para poder llevarlo a cabo. El objetivo principal de esta tesis consiste en estudiar de forma crítica las técnicas habitualmente utilizadas para el análisis de datos en astrosismología, y en particular para el satélite CoRoT, como la transformada de Fourier discreta, y desarrollar técnicas que proporcionen información nueva sobre los datos con el objeto de **identificar la naturaleza de las frecuencias presentes en las curvas de luz observadas por el satélite CoRoT**.

A continuación presentamos el contenido de los distintos capítulos que estructuran el plan de esta tesis:

Parte I: Introducción

- **Capítulo I:** Plan de tesis.

Esquema general del trabajo de tesis describiendo la temática de cada capítulo.

- **Capítulo II:** Marco General

En este capítulo se realiza una descripción breve del fenómeno de las pulsaciones estelares y la clasificación de estas, una revisión de las técnicas de análisis de datos en astrosismología y se presenta la problemática que ha motivado la realización del trabajo de tesis.

Parte II: Eliminación de frecuencias espurias mediante reconstrucción de la señal

- **Capítulo III:** Efectos de la ventana espectral

Consideramos los efectos que tiene el muestreo sobre el espectro de frecuencias de una estrella pulsante al introducir una ventana espectral, también presentamos las técnicas comúnmente utilizadas para evitar esos efectos mediante el relleno de huecos, y proporcionamos evidencias sobre la necesidad de usar nuevas técnicas que no estén limitadas a una base de expresión analítica.

- **Capítulo IV:** Técnicas de relleno de huecos

Presentamos los modelos autoregresivos como técnica para el análisis de series temporales. Desarrollamos una técnica para el relleno de huecos basada en una interpolación mediante procesos autoregresivos de media móvil (ARMA).

- **Capítulo V:** Resultados: series sintéticas y curvas de luz de CoRoT

Mostramos primero la eficiencia del método de interpolación ARMA al ser utilizado sobre simulaciones numéricas de curvas de luz. Posteriormente, se utiliza para reconstruir las curvas de luz de estrellas A-F obtenidas por el satélite

CoRoT. Se contrastan los resultados con los obtenidos mediante otras técnicas y se muestra la rigurosidad de los distintos resultados y la necesidad de hacer uso del método ARMA para eliminar las frecuencias espurias que aparecen en el espectro a causa de la ventana espectral.

Parte III: Estudio de la validez del desarrollo en serie de Fourier de una serie temporal.

- **Capítulo VI:** Consideraciones sobre la consistencia del análisis de frecuencias
Se presenta la condición esencial para el análisis armónico: la analiticidad de la función que representa el observable.

- **Capítulo VII:** El método de las conectividades
Desarrollamos un test para evaluar la validez del desarrollo en series de Fourier de una curva de luz. Este método, basado en los conceptos de continuidad de las derivadas, se prueba con simulaciones de series temporales basadas en el contenido espectral de curvas de luz observadas por CoRoT.

- **Capítulo VIII:** Analiticidad de las series observadas
El método de las conectividades se utiliza para evaluar distintas curvas de luz procedentes de CoRoT, así como de otras misiones espaciales como Kepler y SOHO. Se concluye que la conectividad definida de manera precisa en este trabajo no es una propiedad de los datos suministrados por estos satélites fotométricos ultraprecisos.

- **Conclusiones y Trabajo Futuro**

De la discusión de los resultados obtenidos se alcanzan una serie de hallazgos que configuran las conclusiones más relevantes. A partir de éstas se establecen futuras líneas de trabajo de concreción y explotación de resultados.

Paralelamente, tratamos de exponer un criterio para un análisis de datos riguroso evitando los problemas derivados de la aplicación de los tratamientos que parten de hipótesis ad hoc no verificadas. En definitiva, éste es el error más frecuente en el análisis de datos y puede llevar a una interpretación incorrecta de los fenómenos físicos.

Finalmente, como legado técnico hemos desarrollado una serie de herramientas que serán liberadas a la comunidad astrofísica tras la finalización de esta tesis. Posteriormente, este legado técnico podrá ser incorporado al Observatorio Virtual Español (SVO).

Capítulo 2

Marco general

At first sight it would seem that the deep interior of the Sun and stars is less accesible to scientific investigation than any other region in the universe. Our telescopes may probe farther and farther into the depths of space; but how can we ever obtain certain knowledge of that which is hidden behind substantial barriers? What appliance can pierce through the outer layers of a star a test the condition within?"

Sir Arthur Eddington

2.1 Pulsación Estelar

2.1.1 Introducción a la astrosismología

Pese a que la teoría de la pulsación estelar como tal se ha desarrollado durante el siglo XX las primeras estrellas pulsantes conocidas, las Cefeidas, fueron descubiertas mucho tiempo atrás. John Goodricke descubrió en el invierno de 1784 la variabilidad de δ Cephei pero, de hecho, las primeras estrellas variables conocidas datan del siglo XVI cuando David Fabricius descubrió la estrella Mira ([Christensen-Dalsgaard, 2004](#)). No obstante, pese a la acumulación de hechos observacionales, no fue hasta el trabajo de Eddington ([1918](#)) sobre las Cefeidas, que se sentaron las bases teóricas de la pulsación estelar partiendo de una aproximación adiabática y considerando las estrellas como esferas politrópicas de gas.

Los primeros estudios sobre pulsación estelar se centraron en la comprensión de este fenómeno más que en su utilización para conocer la evolución y estructura estelares. Así, la primera explicación del mecanismo de pulsación estelar se la debemos a Zhevakin ([1953](#)), por identificar el helio ionizado como válvula para generar las pulsaciones en las Cefeidas.

El estudio de las pulsaciones radiales ganó interés por las Cefeidas, pero el estudio de las pulsaciones no radiales, que se conocían por los sistemas estelares binarios desde

2. MARCO GENERAL

hacía mucho tiempo, no experimentó un salto hasta el artículo de Cowling (1941) y no se desarrolló hasta el trabajo de Ledoux (1951), en que se sugirieron éstas como explicación de las variaciones en la anchura de las líneas espectrales de la estrella β Canis Majoris.

A partir de los años 70 se obtuvieron las ecuaciones completas de los interiores estelares y se comenzó a desarrollar el estudio de las oscilaciones no radiales intrínsecas, es decir, no excitadas por binarias sino por mecanismos relacionados con las propiedades internas de las estrellas. De este modo, se estableció una conexión entre pulsaciones y estructura interna de las estrellas que sería el germen de lo que hoy en día se conoce como astrosismología.

La teoría de las pulsaciones no radiales es más compleja que la de las radiales porque, a diferencia de éstas, se deben utilizar tres dimensiones para describir los desplazamientos de la superficie, pero se utilizan una serie de aproximaciones que simplifican el problema. La primera de ellas es considerar la estrella una esfera perfecta de fluido en equilibrio que al ser perturbada experimenta oscilaciones. Éstas pueden ser descritas mediante funciones de coordenadas esféricas (r, θ, ϕ) separables

$$\delta\xi(r, \theta, \phi, t) = \delta\xi(r)Y_\ell^m(\theta, \phi)e^{i\sigma t} \quad (2.1)$$

donde el factor exponencial complejo es la componente armónica en el tiempo y la componente angular $Y_\ell^m(\theta, \phi)$ es a su vez separable en otras dos componentes y viene dada por los armónicos esféricos:

$$Y_\ell^m(\theta, \phi) = P_\ell(\cos \theta)e^{im\phi} \quad (2.2)$$

siendo la parte correspondiente a la coordenada θ el polinomio asociado de Legendre de grado ℓ , el cual representa el número de líneas nodales en la superficie (ver Fig. 2.1). En las ecuaciones anteriores ξ representa la magnitud perturbada, que puede ser la presión o la densidad.

La solución anterior es posible sólo cuando se admite simetría esférica perfecta y cuando además se tiene en cuenta una segunda aproximación importante, que es que las magnitudes perturbadas varíen muy poco en relación a sus valores en equilibrio, de modo que el tratamiento matemático pueda linealizarse. Cuando se usan estas dos aproximaciones junto a la de adiabaticidad en la ecuación de la energía, se obtienen unas ecuaciones de equilibrio estelar (Ecs. de continuidad, movimiento, Poisson y conservación de la energía) que al ser perturbadas dan lugar a la ecuación de pulsación cuyas autofunciones son los armónicos esféricos.

Las observación de las pulsaciones se realiza a través de la medición de las velocidades radiales o a través de la medición de la intensidad luminosa de la estrella. La variación de la intensidad es lo que se conoce como curva de luz estelar y, excepto para

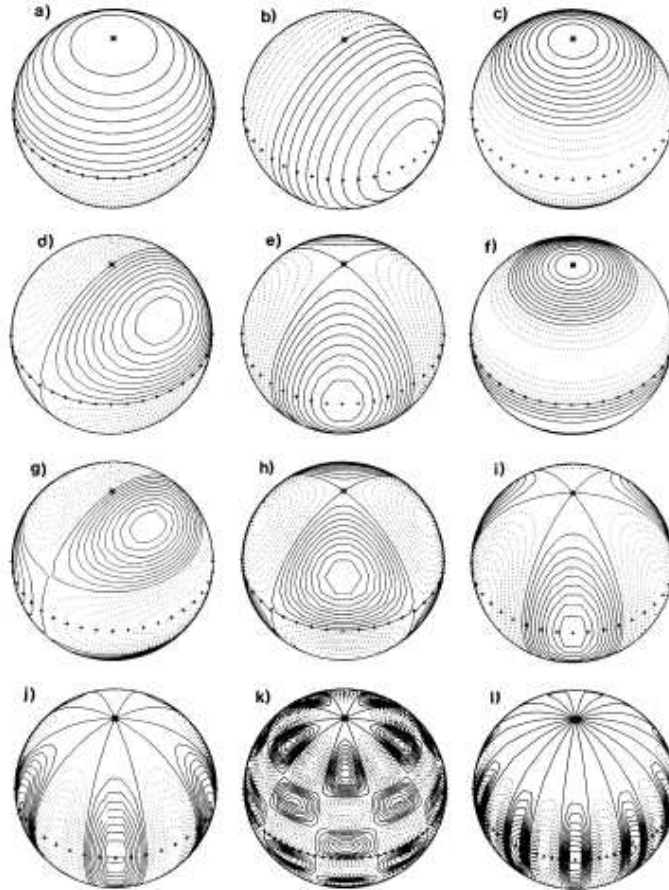


Figura 2.1: Diagrama de contornos de la parte real de los armónicos esféricos. Los contornos positivos se indican por líneas continuas y los contornos negativos por líneas punteadas. El eje de $\theta = 0$ se ha inclinado 45° hacia el observador y viene indicado por un punto. El ecuador se muestra mediante “++++”. Los casos ilustrados se corresponden a: a) $\ell = 1, m = 0$; b) $\ell = 1, m = 1$; c) $\ell = 2, m = 0$; d) $\ell = 2, m = 1$; e) $\ell = 2, m = 2$; f) $\ell = 3, m = 0$; g) $\ell = 3, m = 1$; h) $\ell = 3, m = 2$; i) $\ell = 3, m = 3$; j) $\ell = 5, m = 5$; k) $\ell = 10, m = 5$; l) $\ell = 10, m = 10$. (figura obtenida de [Christensen-Dalsgaard 2003](#))

2. MARCO GENERAL

el caso del Sol en que existe resolución angular suficiente como para obtener medidas sobre el disco, para el resto de las estrellas se mide como luz integrada sobre todo el disco estelar. Así pues, los modelos predicen variaciones en la intensidad luminosa de la estrella que siguen la expresión de una integral sobre todo el área del disco estelar de la función que hemos descrito en la eq. 2.1. No obstante, al observar una curva de luz cualquiera (e.g. Fig.2.2) vemos que la aproximación lineal sólo puede dar una explicación parcial. En el primer panel de la Fig.2.2 puede observarse como la curva de luz está lejos de ser una función armónica. Esto puede atestigüarse también por el proceso de *prewhitening* ilustrado en los demás paneles de la figura. Este proceso consiste en un ajuste iterativo de mínimos cuadrados de funciones sinusoidales. En cada iteración se elimina de la serie la función ajustada de modo que, la serie resultante, que se conoce como “residuos”, se utiliza como entrada para la siguiente iteración. Así pues, si la señal es una función armónica pura, tras un número de iteraciones el algoritmo converge a una serie compuesta únicamente por ruido blanco gaussiano. En cambio, en la figura se observa como, pese a que la amplitud se va reduciendo en los distintos pasos del *prewhitening*, los residuos finales tras eliminar 1686 frecuencias siguen mostrando variaciones no aleatorias. Este es un fenómeno que no está explicado a día de hoy.

Pese a la dificultad de interpretación de los espectros de frecuencias de las estrellas pulsantes, la astrosismología ha sido muy fructífera como herramienta para el conocimiento de los interiores estelares y propiedades de las estrellas como su edad o su radio. Una importante aplicación en la que ha demostrado su potencia ha sido en el caso de las estrellas enanas blancas pulsantes. Hasta mediados de los años 70 se desconocía la existencia de estas estrellas. Los primeros resultados se obtuvieron para las enanas blancas DA, caracterizadas por la presencia de hidrógeno en sus espectros, y con temperaturas efectivas en torno a 10000 K. Desde entonces se han hallado otros grupos de enanas blancas pulsantes, como las variables DB o las de nebulosas planetarias (PNNV), que se hallan en un estado de evolución rápida. Éstas estrellas, en especial las variables DA, experimentan cambios en el periodo de pulsación debido a su evolución a lo largo de la secuencia de las enanas blancas. Así pues, las mediciones de distribución en luminosidad de estas estrellas proporcionan una de las mejores estimaciones que existen sobre la edad de la Galaxia.

Otro logro importante de la astrosismología ha sido la aplicación con éxito de la aproximación WKBJ, muy utilizada en Mecánica Cuántica (Shibahashi, 1979) logrando obtener una expresión analítica de las autofrecuencias de modos no radiales de alto orden n (Tassoul, 1980). Esta expresión (cf. Sec. 5.1.1.3) permite determinar con precisión propiedades como la densidad de la estrella. La aproximación asintótica es

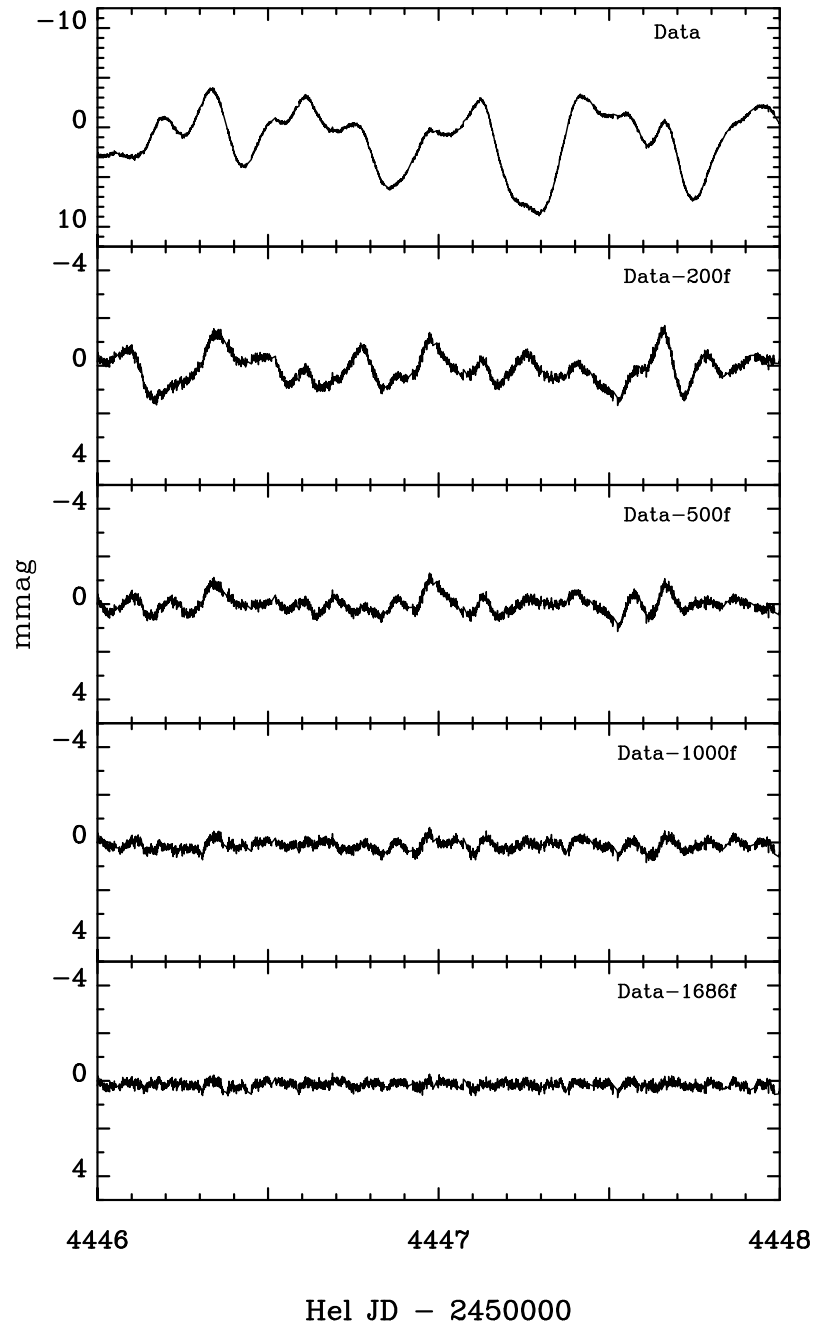


Figura 2.2: Curva de luz de la estrella HD 49434 y residuos tras eliminar distintos conjuntos de frecuencias ajustadas. (figura obtenida de [Chapellier et al. 2011](#))

2. MARCO GENERAL

especialmente útil en estrellas de tipo solar ya que estas tienen excitados modos p^1 de orden muy alto. El descubrimiento de las oscilaciones solares de 5 minutos de periodo por Leighton y Noyes en 1962 fue otro de los hallazgos más notables en el estudio de las pulsaciones estelares dando lugar a una rama de la astrosismología conocida como Heliosismología.

Por otro lado, se han desarrollado modelos teóricos más sofisticados que han ganado en la precisión de las predicciones al incluir otros fenómenos en las ecuaciones como son: la rotación (Suárez et al., 2006), la no adiabaticidad debido a la interacción de las pulsaciones con la atmósfera (Moya and Garrido, 2008), la interacción pulsación-convección (Grigahcène et al., 2005), etc.

2.1.2 Clasificación de estrellas pulsantes

En la tabla 2.1 encontramos los principales tipos de pulsación estelar hallados. Cabe observar que las gigantes rojas aparecen en el mismo grupo que las estrellas de tipo solar porque presentan los mismos patrones en el espectro de frecuencias que éstas. Esto es, un espectro caracterizado por picos uniformemente espaciados y una distribución de potencia gaussiana que cubre todo un extenso rango de frecuencias excitadas.

Si se tienen en cuenta los rangos de temperaturas efectivas y luminosidades relativas de las estrellas δ Sct, RR Lyrae y Cefeidas, se observa que estos grupos de estrellas caen todos en una misma banda conocida como banda de inestabilidad de las Cefeidas (o Clásica), ya que éstas fueron las primeras en ser descubiertas. La localización de estos grupos de estrellas en esta banda es indicativo de que comparten el mismo mecanismo de excitación, el mecanismo de opacidad asociado a la segunda zona de ionización del helio. Otras estrellas que también tienen sus oscilaciones excitadas por éste mecanismo son las SPB (*Slowly Pulsating B*) y las β Cephei, pero estas no están situadas en la banda de inestabilidad clásica y el mecanismo de opacidad está relacionado con los elementos del grupo del hierro. Por otra parte, las estrellas roAp (*rapidly oscillating A peculiar*) sí se hallan en la banda de inestabilidad pero sus oscilaciones están ligadas a los grandes campos magnéticos que poseen.

Las γ Dor se encuentran en la intersección entre la secuencia principal y el borde rojo de la banda de inestabilidad y son pulsantes de largo periodo, es decir, por lo general modos g, pero su mecanismo de excitación no es del todo conocido aún. Algunas de estas estrellas presentan características espectrales híbridas con las estrellas δ Sct, como es el caso de la estrella KIC 006187665 observada por *Kepler* que estudiamos en la parte III de este trabajo.

¹Oscilaciones en que la fuerza restauradora es la presión, a diferencia de los modos g u ondas gravitatorias, en que la fuerza restauradora es la gravedad.

Por último, se conocen pocos candidatos de estrellas pulsantes pre-secuencia principal (pre-SP en la tabla) pero en alrededor de 30 de ellos se ha sugerido la existencia de oscilaciones de tipo δ Sct.

Tipo de pulsación	Tipo de modos	Periodos	Amplitudes	$\log T_{eff}$ (K)	$\log L/L_{\odot}$
Solar-like	p	3-10 min	\sim ppm	[3.70, 3.82]	[-0.5, 1.0]
Gigantes rojas	p	\sim horas	\sim 10 ppm	[3.65, 3.70]	[-0.5, 2.0]
γ Dor	g	8 h a 5 d	<50 mmag	[3.83, 3.90]	[0.7, 1.1]
δ Sct	p	15 min a 8 h	<0.3 mag	[3.82, 3.95]	[0.6, 2.0]
roAp	p	5-22 min	<10 mmag	[3.82, 3.93]	[0.8, 1.5]
SPB	p	15 min a 8 h	<50 mmag	[4.05, 4.35]	[2.0, 4.0]
β Cephei	p, g	1-12 h (p); días (g)	<0.1 mag (p); <0.01 mag (g)	[4.25, 4.50]	[3.2, 5.0]
Be pulsante	p, g	0.1-5 d	<20 mmag	[4.05, 4.50]	[2.0, 5.0]
pre-SP: T Tauri	p,g	1-8 h (p); 8h a 5d (g)	<5 mmag	-	-
pre-SP: Herbig Ae/Be	g	1-8 h	<5 mmag	-	-
sdB	p,g	80-800 s (p); 0.5-3 h	<0.1 mag (p); <0.01 mag (g)	[4.20, 4.50]; [4.40, 4.60]	[1.2, 2.2]; [1.2, 2.6]
PNNV	g	5 h a 5 d	<0.3 mag	-	-
GW Vir	g	5-80 min	<0.2 mag	[4.80, 5.10]	[1.5, 3.5]
DBV	g	2-16 min	<0.2 mag	[4.40, 4.60]	[-1.0, 0.7]
DAV	g	0.5-25 min	<0.2 mag	[3.95, 4.15]	[-2.6, -2.2]
RR Lyrae: RRab	F	~ 0.5 d	<1.5 mag	[3.78, 3.88]	[1.4, 1.7]
RR Lyrae: RRc	FO	~ 0.3 d	<0.5 mag		
RR Lyrae: RRd	F+FO	0.3-0.5 d	<0.2 mag		
Cefeidas Tipo II: W Vir	F	10-30 d	<1 mag	[3.70, 3.90]	[2.0, 4.0]
Cefeidas Tipo II: BL Her		1-5 d			
RV Tauri	F	30-150 d	<3 mag	[3.60, 3.90]	[3.2, 4.2]
Cefeidas Tipo I	F	1-50 d	<1 mag	[3.55, 3.85]	[2.0, 5.5]
Mira: SRa, SRb	$\ell = 0$	>80 d	<8 mag	[3.45, 3.75]	[2.5, 4.0]
Mira: SRc	$\ell = 0$	>80 d	<1 mag		
Mira: SRd	$\ell = 0$	<80 d	<1 mag		

Tabla 2.1: Clasificación de estrellas pulsantes. Nótese que los intervalos son aproximados.

2.1.3 astrosismología desde satélites

- WIRE (Wide-Field Infrared Explorer): satélite lanzado en 1999, su objetivo era hacer un survey del universo infrarrojo pero no pudo completar sus objetivos científicos de modo que los sistemas fueron redirigidos para la observación de estrellas brillantes en el marco de la astrosismología y la búsqueda de exoplan-

2. MARCO GENERAL

etas. Los primeros resultados en astrosismología desde el espacio fueron sobre la gigante roja α UMa (Buzasi et al., 2000).

- MOST (Microvariability and Oscillations in STars): Primer satélite cuyo objetivo principal fue el estudio de la astrosismología. Lanzado en 2003, iba equipado de un pequeño telescopio con un espejo de 15 cm y una cámara de luz blanca. Sus principales objetivos fueron las oscilaciones de tipo solar, exoplanetas orbitando estrellas de tipo solar, y el estudio de estrellas Wolf-Rayet (estrellas masivas evolucionadas). Inicialmente, la no detección de oscilaciones en Procyon provocó una controversia en la comunidad científica (Matthews et al., 2004).
- CoRoT (CONvection, ROTation and planetary Transits): fue lanzado a finales de 2006. Consiste en un telescopio de 27 cm con cuatro CCDs capaz de detectar oscilaciones con una precisión de hasta $1 \mu\text{mag}$, que es mucho mayor que desde tierra. Su órbita polar le permite tener un campo de visión no perturbado por el Sol hasta 150 días continuos (*Long Run*). Tras eso el satélite rota para observar en un campo intermedio durante un periodo de unos 20 días (*Short Run*) y a continuación rota en dirección antisolar para observar durante otros 150 días de observación. CoRoT utiliza dos de sus CCDs con tiempo de integración de 32 s para la observación de las oscilaciones estelares y las otras dos CCDs con tiempo de integración de 512 s para la detección de exoplanetas (Auvergne et al., 2009). Actualmente CoRoT está en fase de decomisionado por fallo en uno de los ordenadores de a bordo. No obstante, el legado de CoRoT es enorme, tanto en la detección de exoplanetas con decenas de descubrimientos, como en astrosismología con la primera detección de oscilaciones no radiales en gigantes rojas (De Ridder et al., 2009).
- *Kepler*: Lanzado por la NASA a principios de 2009 con el objetivo principal de descubrir planetas similares a la Tierra orbitando otras estrellas y estimar cuántas estrellas de ese tipo hay en nuestra galaxia (Gilliland et al., 2010). El espejo primario del telescopio de *Kepler* es de 1.4 metros y el plano focal consta con 42 CCDs que dan una resolución de 95 Mpx. Las CCDs tienen dos cadencias de integración, 58.89 s y 29.4 min. Las características de esta misión la hacen ideal para estudiar astrosismología además de exoplanetas, por ello los datos recogidos se utilizan también para medir las oscilaciones estelares, particularmente, para las de tipo solar. Actualmente cuenta en su haber casi un millar de detecciones de exoplanetas. Pese a que la misión ha sido prolongada hasta 2016 desde mediados de 2013 está funcionando con dos de los cuatro giróscopos con que contaba inicialmente por lo que el sistema de apuntado ha dejado de funcionar de forma correcta.

- PLATO (PLANetary Transits and Oscillations of stars): Es una misión de clase M aprobada por la Agencia Espacial Europea para ser lanzada en 2024 en una órbita en torno al punto de Lagrange L2. El objetivo principal de esta misión es descubrir y caracterizar sistemas exoplanetarios cercanos con una precisión en la determinación de la masa de los planetas de hasta el 10 %, de hasta 2 % del radio de los planetas, y de hasta el 10 % de la edad estelar. PLATO detectará planetas del tamaño de la Tierra y Super-Tierras en la zona habitable alrededor de estrellas de tipo solar, medirá las oscilaciones de tipo solar en las estrellas que alberguen los exoplanetas, y medirá oscilaciones de estrellas pulsantes clásicas. Así pues PLATO se plantea como la misión sucesora de CoRoT y *Kepler*.

2.2 Técnicas de análisis de series temporales en astrosismología

La astrosismología es el estudio de la estructura interna de las estrellas y su evolución a través de las pulsaciones. Las pulsaciones dan lugar a perturbaciones en la densidad que se traducen en variaciones del brillo u otras magnitudes observables. El registro de estas magnitudes en el tiempo permite computar sus variaciones y analizando la serie así generada, determinar las frecuencias de las pulsaciones de la estrella. Éstas, salvo casos excepcionales como es el de las Cefeidas, no suelen aparecer como variaciones sinusoidales sencillas. Las pulsaciones suelen tener una componente no lineal que hace que en el espectro de frecuencias aparezcan no sólo las frecuencias correspondientes a los modos de pulsación sino también sus armónicos. Si la serie temporal no se ha muestreado de forma continua durante toda la observación, los picos se verán convolucionados con una función conocida como ventana espectral (ver Cap. 3) que está relacionada con el muestreo de los datos. Ésto da lugar a que aparezcan otros picos espurios que no se corresponden a los modos de pulsación. Además, el ruido fotónico, a pesar de tener una distribución espectral de ruido blanco producirá siempre picos en el espectro de frecuencias de mayor o menor amplitud debido a que la observación es finita y el espectro no es completamente plano. Si a todo esto añadimos la posibilidad de interacción entre modos, los efectos de amortiguamiento producidos por la convección, variaciones producidas por ciclos magnéticos, granulación, flares y otros artefactos, la detección de frecuencias propias de los modos de pulsación se convierte en una disciplina digna del mejor detective.

Así pues, existe toda una pléthora de técnicas para el análisis de las series temporales que se obtienen al observar estrellas pulsantes según las condiciones de la observación y el tipo de variabilidad. En la tabla 2.2 recogemos las más comunes y más abajo comentamos su dominio y limitaciones.

2. MARCO GENERAL

Métodos basados en la transformada de Fourier	
Periodograma Lomb-Scargle	Scargle (1982)
Periodograma Lomb-Scargle Generalizado	Zechmeister and Kürster (2009)
Significancia Espectral (SigSpec)	Reegen (2007)
Algoritmo CLEAN	Högbom (1974)
Periodograma multiarmónico	Schwarzenberg-Czerny (1996)
Maximum Likelihood Estimation (MLE)	Appourchaux et al. (1998)
Métodos que no usan una base armónica	
Transformada Wavelet	Régulo et al. (2005)
Minimización de la dispersión de las fases (PDM)	Stellingwerf (1978)
Algoritmo Box-fitting	Kovács et al. (2002)
Métodos basados en el dominio del tiempo	
Método de Entropía Máxima (MEM)	Burg (1975)
Método de la Entropía de la Información	Cincotta et al. (1995)

Tabla 2.2: Técnicas para el análisis de series temporales.

Para hallar las frecuencias de pulsación se ajusta un modelo a las observaciones que será más o menos complejo según sea el caso. Así, el modelo más básico de una observación Y_n en tiempo t_n consiste en una descomposición en suma de senos que puede escribirse de forma lineal como:

$$Y_n(t_n) = \mu + \sum_i A_i X_{i,n} + \epsilon_n \quad \text{donde} \quad X_{i,n} = \sin[2\pi(f_i t_n + \phi_i)] \quad (2.3)$$

donde μ es la media y ϵ_n es la componente de ruido blanco, gaussiano e independiente. Este modelo puede ajustarse con cualquiera de los métodos existentes para el ajuste de mínimos cuadrados. Para hacer el ajuste es necesario hacer previamente una estimación de las frecuencias que ha de hacerse mediante técnicas de análisis de Fourier. La técnica más extendida en astrosismología para esto es la introducida por [Deeming \(1975\)](#) y generalizada por [Lomb \(1976\)](#) y Scargle. Según ésta la determinación de las frecuencias puede hacerse mediante un estimador de la densidad espectral conocido como periodograma. Éste viene dado por:

$$P(f) = \frac{1}{2} \left[\frac{[\sum_j Y_j \cos(2\pi f t'_j)]^2}{\sum_j \cos^2(2\pi f t'_j)} + \frac{[\sum_j Y_j \sin(2\pi f t'_j)]^2}{\sum_j \sin^2(2\pi f t'_j)} \right] \quad (2.4)$$

donde $t'_j = t_j - \tau$ es una redefinición del parámetro t que se realiza para que el comportamiento estadístico sea consistente incluso cuando el muestreo de la serie

observada es no uniforme. El parámetro τ viene dado por:

$$\tan (2\pi f\tau) = \frac{\sum_j \sin (2\pi f t_j)}{\sum_j \cos (2\pi f t_j)} \quad (2.5)$$

Al representar la $P(f)$ las componentes armónicas aparecen identificadas como picos correspondiendo a funciones sinc ($\text{sinc } x = \frac{\sin x}{x}$) de anchura finita que se aproximan a una delta de Dirac cuando el tiempo de observación $t_{\text{obs}} \rightarrow \infty$. No obstante, cuando el muestreo es no-uniforme y altamente irregular pueden aparecer en el periodograma picos de una amplitud considerable no debidos a la señal. Estos son lo que se conoce comúnmente como alias producidos por la ventana de observación. La existencia de alias en el periodograma implica que la frecuencia de mayor amplitud de la señal podría no estar localizada donde se encuentra el pico de mayor amplitud. Para eliminar los alias que corresponden a la convolución de la ventana de observación con cada una de las frecuencias de la señal se suele hacer uso de un procedimiento conocido como *prewhitening*. Este procedimiento consta de cuatro pasos:

- Identificación de la frecuencia de mayor amplitud f_i en el periodograma: para ello el periodograma se calcula con un sobremuestreo que permite diferenciar picos que, de otra manera, serían indistinguibles.
- Precisión: se recalcula el periodograma sobre un rango de frecuencias en torno a la f_i con un sobremuestreo mayor para obtener una mejor medida de la frecuencia.
- Ajuste: usando todas las frecuencias identificadas previamente se realiza un ajuste sobre las amplitudes y fases.
- Prewhitening: se elimina la señal ajustada en el paso anterior de los datos y se itera desde el principio para hallar la siguiente frecuencia.

Este proceso se repite mientras se cumpla un determinado criterio de significancia basado en las amplitudes de las frecuencias identificadas en el primer paso. El ajuste mencionado en el esquema anterior es un ajuste lineal que fue propuesto por [Vaníček \(1971\)](#). Algunos códigos más recientes contienen variaciones sobre este esquema, como PERIOD04 ([Lenz and Breger, 2005](#)), que adopta un esquema más robusto que el basado en un algoritmo de optimización no lineal, el algoritmo de Levenberg-Marquadt ([Press et al., 1992](#)). Cuando la señal contiene pocas frecuencias esta aproximación permite recuperar los valores de frecuencias y amplitudes incluso aunque éstas estén muy cerca ($\Delta f \ll 1/T$). Sin embargo, cuando el número de parámetros a ajustar es alto, incluso un algoritmo de optimización no lineal puede no converger a los valores apropiados. En dichos casos, los valores correctos se obtienen sólo si las

2. MARCO GENERAL

frecuencias son moderadamente cercanas ($\Delta f > 0.5/T$), de lo contrario el comportamiento del ajuste no lineal puede ser muy impredecible.

Además de las limitaciones mencionadas a este método existen otras por la propia naturaleza de la serie a estudiar. De este modo existen variantes para tratar de solucionar los problemas encontrados en cada situación. Así, por ejemplo, se usa el periodograma de Lomb-Scargle generalizado en series con pocos puntos en las que la media real es muy distinta a la encontrada en el conjunto de datos disponible; el algoritmo CLEAN se utiliza cuando los efectos del muestreo hacen muy complicado el análisis, que es el caso de las observaciones de radiointerferometría; cuando la señal tiene variaciones periódicas no-sinusoidales el periodograma multiarmónico de Schwarzenberg-Czerny concentra en un sólo pico la potencia que aparecería dispersa en un conjunto de picos en un periodograma de Lomb-Scargle; cuando la señal consiste en pulsaciones excitadas estocásticamente y amortiguadas como ocurre en las estrellas de tipo solar, los picos correspondientes a cada componente en frecuencias se ensanchan y para determinar los parámetros de forma correcta es necesario hacer un ajuste de perfiles lorentzianos mediante el método de *Maximum Likelihood Estimation*. También existen otras aproximaciones que no se basan en una transformada de Fourier sino en otras bases como es el caso de la Transformada Wavelet, que es particularmente apropiada para el estudio de señales no-estacionarias; también es el caso del algoritmo Box-Fitting, basado en funciones caja, que se usa para la detección de tránsitos de planetas en las curvas de luz. Por otro lado, el algoritmo de Stellingwerf no hace uso de ninguna función base sino que busca periodicidades en la serie mediante minimización de la dispersión de las fases. Por último, también se encuentran en la literatura otros métodos que estudian las propiedades de la serie en el dominio del tiempo sin llevar a cabo ninguna transformación de dominio, como son el Método de Máxima Entropía de Burg, y el Método de Entropía de la Información de Cincotta.

2.3 Criterio de corrección científica

Existe una enorme diversidad de técnicas para el análisis de series temporales, algunas de las cuáles hemos visto en el apartado anterior, y cientos más que hemos dejado en la literatura. ¿Cómo elegir la técnica adecuada?, ¿cómo se pueden combinar las distintas técnicas?, ¿qué criterio se ha de seguir a la hora de hacer el análisis?, etc. Para finalizar este capítulo introductorio de la memoria formulamos unos principios que nos sirven para responder estas preguntas y tener una referencia constante en el modo de proceder durante el análisis de las series temporales.

Siguiendo los principios del método científico, el criterio de corrección científica se basa en introducir el mínimo de hipótesis cuando se hace un análisis observacional. Este criterio tiene menor relevancia cuando lo que se está haciendo es un modelo teórico ya que las hipótesis no introducen a priori ningún sesgo en los resultados y la validez de éstas viene justificada por la propia capacidad de descripción de las hipótesis. En cambio, cuando lo que se realiza es un análisis observacional es crucial evitar rigurosamente el factor humano y presentar el análisis de la forma más objetiva posible.

Siguiendo este criterio establecemos ahora unas pautas o postulados para el análisis de datos que obligan a la rigurosidad matemática y son ineludibles para respetar la objetividad científica:

1. **Las hipótesis bajo las cuales se realiza el análisis de datos deben partir de bases físicas y no de una aproximación matemática.** El análisis de datos es un método experimental de la física y como tal, sus bases deben tratarse desde un punto de vista teórico y no meramente como una técnica computacional. Asimismo ocurre con la interpretación de los resultados.
2. **La forma funcional usada para la identificación de cualquier patrón en los datos debe ser lo más general posible y los grados de libertad deben estar restringidos únicamente por la propia configuración del sistema.** El proceso de análisis de datos consiste en identificar un patrón en los datos analizados hallando su representación matemática, y a continuación tratar de encontrar la causa del patrón en términos de un modelo físico adecuado. De proceder a la inversa, es decir, determinar el patrón inherente a los datos y sus propiedades matemáticas en base a un modelo físico supuesto a priori se corre el riesgo de incurrir en un sesgo.
3. **El análisis de datos debe ser completamente objetivo y no puede depender de definiciones circunstanciales o ambiguas.** La definición de lo que es ruido y lo que es señal debe ser una definición físico-matemática. En consecuencia se derivan definiciones en los dos puntos siguientes.
4. **Definimos como ruido la componente de una serie que se corresponde físicamente con un error de medida, y matemáticamente a una función estocástica cuyos valores son estadísticamente independientes y cuya distribución es una gaussiana.**
5. **Toda componente de una serie que no se ajuste a la definición de ruido dada, es señal y debe ser estudiada como tal, independientemente de**

que la información que contiene sea de mayor o menor relevancia para una interpretación de los datos.

6. **El espacio cubierto en el análisis de datos debe ser máximo con el mínimo de hipótesis.** El análisis de datos parte siempre de una representación de éstos de forma que su información pueda ser interpretada. Toda representación se basa en una serie de hipótesis que definen el tipo de información que se puede extraer. En cada caso el conjunto de hipótesis mínima respecto a la información disponible es el más objetivo. Sólo en este caso se pueden evaluar los resultados de forma consistente sin tener que evaluar además las hipótesis en que se basa la representación.
7. **Un método adecuado para el análisis de datos debe ser suficientemente simple como para que sus propiedades estadísticas puedan ser determinadas con facilidad.** De lo contrario, el método se convierte en un simple procedimiento matemático sin una justificación física para su realización. En este sentido, consideramos que solamente son definibles aquellas propiedades cuya distribución es gaussiana.
8. **Las técnicas de análisis utilizadas deben ser consistentes con las características de la serie y los parámetros de la observación.** Los métodos que se utilizan para computar el espectro de potencias (métodos de Fourier) sufren las consecuencias de suposiciones no realistas sobre la extensión de los datos fuera de intervalo conocido. Estas suposiciones llevan a ventanas espectrales que limitan severamente la resolución de los espectros de observaciones cortas. La aproximación MEM propuesta por Burg (equivalente a procesos ARMA) proporciona una estimación espectral que carece de las limitaciones de la DFT en las condiciones críticas mencionadas.
9. **La elección de unidades debe obedecer a razones físicas.** Cualquier transformación sobre éstas debe estar ligada a la invariancia en la determinación de las propiedades estadísticas, y no a cuestiones estéticas o históricas. Así pues, se puede demostrar que una transformación de unidades sobre una función no lineal puede causar que propiedades estadísticas como la media y la varianza no sean consistentes. En este sentido, el cambio de intensidades I_n a magnitudes m_n dado por

$$m_n = -2.5 \log(I_n/I_0)$$

que es una transformación no lineal, no se utiliza en el contexto de este trabajo. Así mismo evitamos también la habitual conversión a partes por millón (ppm)

$$I = 10^6 \left(\frac{\mu_I}{I} - 1 \right)$$

que, además de ser una transformación no lineal, puede tener singularidades.

2. MARCO GENERAL

Parte II

Eliminación de frecuencias espurias mediante reconstrucción de la señal



Cuando se analizan series temporales es corriente encontrar huecos ya sea debido a un muestreo no uniforme de la serie o a errores instrumentales que obligan a descartar algunas medidas. Los huecos suponen un handicap cuando se trata de identificar frecuencias de pulsación en las curvas de luz a través de técnicas de Fourier.

El procedimiento habitual para evitar el efecto que producen dichos huecos sobre el análisis de frecuencias suele consistir en una de estas dos soluciones: estudiar la ventana espectral o realizar un relleno de huecos. En el primer caso se analiza la función definida por el propio muestreo de la serie conocida como función ventana, o más bien su representación en frecuencias, la ventana espectral, que convoluciona la representación de la señal en frecuencias. Esta solución, quizás la más extendida, no está falta de dificultades ya que, en la práctica la función ventana suele ser bastante complicada y no es posible deconvolucionar la señal de forma que la decisión sobre qué frecuencias son reales y cuáles son debido a la convolución de la ventana espectral acaba siendo heurística. La segunda solución, realizar un relleno mediante una interpolación, suele reducirse a una interpolación lineal en el caso de huecos de pocos datos. Sin embargo, en la mayoría de los casos en que se rellenan huecos haciendo ajustes lineales se está haciendo algo que es a la vez erróneo e innecesario. Los casos en los que una serie sigue una relación lineal son extraordinariamente escasos y en la práctica las relaciones que describen de forma correcta los datos no son ni siquiera aproximadamente lineales. Casi todas las series resultan ser no-lineales de modo que con un ajuste lineal el investigador incurre en un error substancial que puede generar inconsistencias en el análisis de la serie. Incluso si en lugar de hacerse un ajuste lineal se realiza un ajuste con cualquier otra función de forma cerrada sin tener una razón teórica bien fundamentada para poder asegurar que los datos se generen mediante ese tipo de relación, se incurrirá en una inconsistencia en el análisis. Por lo general la distribución de los datos en una serie temporal suele ser mucho más rica que lo que cualquier función de forma cerrada pueda ajustar.

En esta parte de la memoria exponemos un trabajo en el que se ha desarrollado un método para rellenar huecos ajustando las series sin hacer uso de ningún modelo basado en una función de forma cerrada. Con esto se evita la preconcepción inherente a todo modelo ad-hoc y de esta forma el método es capaz de recuperar la estadística intrínseca a los datos de una forma natural, es decir, usando sólo la información propia de los datos y sin introducir ninguna señal artificial.

Comenzamos la exposición en el Cap. 3 definiendo la ventana espectral y profundizando sobre los efectos espúreos que pueden aparecer en la detección de frecuencias como consecuencia del muestreo realizado. Revisamos las soluciones propuestas en la literatura para evitar esos efectos y estudiamos entre estas los distintos métodos de relleno de huecos que se han desarrollado.

En el Cap. 4 exponemos el método de interpolación basado en modelos ARMA (Pascual-Granado et al., 2012) que hemos desarrollado dentro de este trabajo de tesis. Este método se evalúa estadísticamente y es puesto a prueba con series sintéticas que simulan las variaciones típicas observadas en las curvas de luz de estrellas pulsantes.

Finalmente, la eficiencia del método se muestra al aplicar este sobre datos reales procedentes del satélite CoRoT tanto para el caso de curvas del campo de sismología estelar como para el de exoplanetas. Estos resultados se presentan en el Cap. 5.2.

Capítulo 3

Efectos del muestreo en la detección de señal: la ventana espectral

“All of physics is either impossible or trivial. It is impossible until you understand it, and then it becomes trivial.”

Ernest Rutherford

Shannon, en su famoso artículo en el que se sientan las bases del análisis de señal (Shannon, 1949), establece que una función $f(t)$ cuyas frecuencias están limitadas por $\Delta\nu$ queda completamente determinada por los valores $f(t_n)$ con $n = -\infty \dots +\infty$ que toma la función en una serie de puntos t_n que están espaciados por $\Delta t = 1/2\Delta\nu$. Se demuestra (Appourchaux, 2011), además, que la representación de $f(t)$ dada a través de esos valores de la función es

$$f(t) = \sum_{n=-\infty}^{n=+\infty} f(t_n) \operatorname{sinc}\left(\frac{t-t_n}{\Delta t}\right)$$

Así, pese a que la reconstrucción exacta de la función requeriría una suma infinita y en la práctica las series temporales consisten sólo en un número finito de medidas de una función, se puede apreciar aquí la importancia de contar con un muestreo uniforme para analizar la función.

En la expresión anterior puede constatarse que los valores de la función $f(t_n)$ constituyen los coeficientes de su desarrollo. Por otra parte, estos permiten a su vez expresar la función en el espacio de frecuencias a través de la descomposición de Fourier:

$$F(\nu) = \Delta t \sum_{n=-\infty}^{n=+\infty} f(t_n) e^{i2\pi\nu t_n} \quad (3.1)$$

Esta representación es extremadamente útil cuando se analizan señales compuestas por un conjunto de funciones armónicas caracterizadas por las frecuencias ν_1, ν_2, \dots

3. EFECTOS DEL MUESTREO EN LA DETECCIÓN DE SEÑAL: LA VENTANA ESPECTRAL

ya que la $F(\nu)$ solamente toma valores distintos de cero alrededor de esas frecuencias, siguiendo una delta de Dirac.

En la práctica debido a que las series además de ser discretas son finitas, la sumatoria en la Eq. 3.1 termina en el término N . Esto es lo que conocemos como Transformada de Fourier Discreta (TFD), que no es una representación exacta sino una aproximación que presenta una serie de diferencias sobre la descomposición en serie de Fourier. Nos ocupamos ahora de estas diferencias.

Esencialmente las diferencias entre la descomposición en serie de Fourier de una función y su TFD vienen dadas por interferencias entre las frecuencias, ya sean estas cercanas o distantes. La interferencia entre frecuencias cercanas es consecuencia de la anchura de los picos debido a que la serie es finita. En el segundo caso, la interferencia entre frecuencias distantes ocurre por el submuestreo de la señal que da lugar a lo que se conoce como aliasing. Para un muestreo arbitrario los 2 tipos de interferencias mencionados no se pueden separar y dan lugar a lo que se conoce como ventana espectral.

Un importante resultado matemático en este campo establece que la convolución de la transformada de Fourier real y la función que describe la ventana espectral coincide con la TFD (Deeming, 1975). Esto es

$$\frac{1}{N} F_N(\nu) = F(\nu) * V_N(\nu) \quad (3.2)$$

siendo $V_N(\nu)$, la ventana espectral, una función que se obtiene a través de los tiempos de observación de la función como

$$V_N(\nu) = \frac{1}{N} \sum_{n=1}^N e^{i2\pi\nu t_n} \quad (3.3)$$

El factor $1/N$ en las expresiones anteriores es simplemente una normalización para que $V_N(0) = 1$. La Eq. 3.2 define de forma correcta la TFD que, pese a ser tratada como tal no puede ser considerada una aproximación numérica de la transformada de Fourier de la función y es, de hecho, dimensionalmente distinta a ésta. Una consecuencia importante de esta definición es que, aunque la TFD sea una función distinta a la transformada de Fourier su existencia depende de esta y, por tanto, se deben comprobar las condiciones de existencia, e.g. si la función $f(t)$ es absolutamente integrable. Para comprobar esto es necesario hacer una correcta reconstrucción de la señal, es decir, sin utilizar funciones de forma cerrada para ello.

Por otra parte, para un muestreo uniforme y N muy grande la ventana espectral se reduce a una función peine de Dirac lo que simplifica enormemente el análisis. Ésto muestra nuevamente la ventaja de obtener la serie muestreada de forma constante.

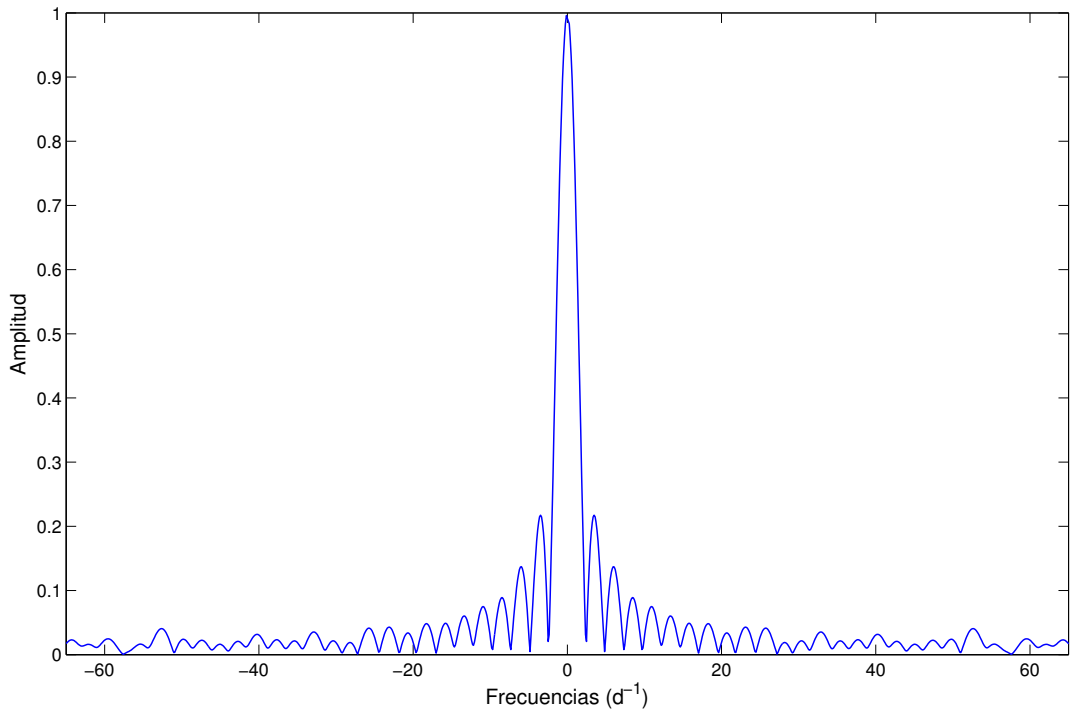


Figura 3.1: Ventana espectral de una serie de velocidades radiales tomadas desde tierra. Nótese los lóbulos debido al periodo de observación. La serie está constituida por 57 medidas tomadas desde el observatorio de Calar Alto durante una noche de observación. El espectro mostrado se ha sobremuestreado 20 veces para mostrar el detalle de los lóbulos alrededor del pico en $\nu = 0$.

Como puede verse en la Fig. 3.1, una ventana espectral típica presenta un pico central con una anchura en torno a $1/T$, que determina la interferencia entre frecuencias cercanas, y una serie de picos colindantes de menor amplitud que surgen debido a las irregularidades del muestreo. Sobre estas es sobre las que estamos interesados ya que el relleno de huecos recupera la uniformidad en el muestreo y elimina esta contribución espuria que, pese a ser de poca amplitud, en el caso de las series temporales tomadas por satélites, las interferencias afectan a todo el espectro.

En la Fig. 3.2 mostramos otro ejemplo de ventana espectral, en este caso de una curva de luz observada por el satélite *Kepler*. La ventana tiene un hueco de 2.19 días debido a la entrada de la electrónica en modo seguro, y se han eliminado 557 puntos identificados como outliers según las correcciones instrumentales de [García et al. \(2011\)](#). En este caso podemos observar un pico central mucho más estrecho gracias al periodo de observación de un mes. Los lóbulos se reducen también considerablemente y aunque podemos ver que la ventana espectral dista de ser una delta de Dirac debido a los numerosos picos que aparecen, éstos son de baja amplitud y como mucho podrían repercutir en que el ruido de fondo al analizar el espectro de la estrella aumentara pero no es de esperar que interfieran con las frecuencias significativas.

3. EFECTOS DEL MUESTREO EN LA DETECCIÓN DE SEÑAL: LA VENTANA ESPECTRAL

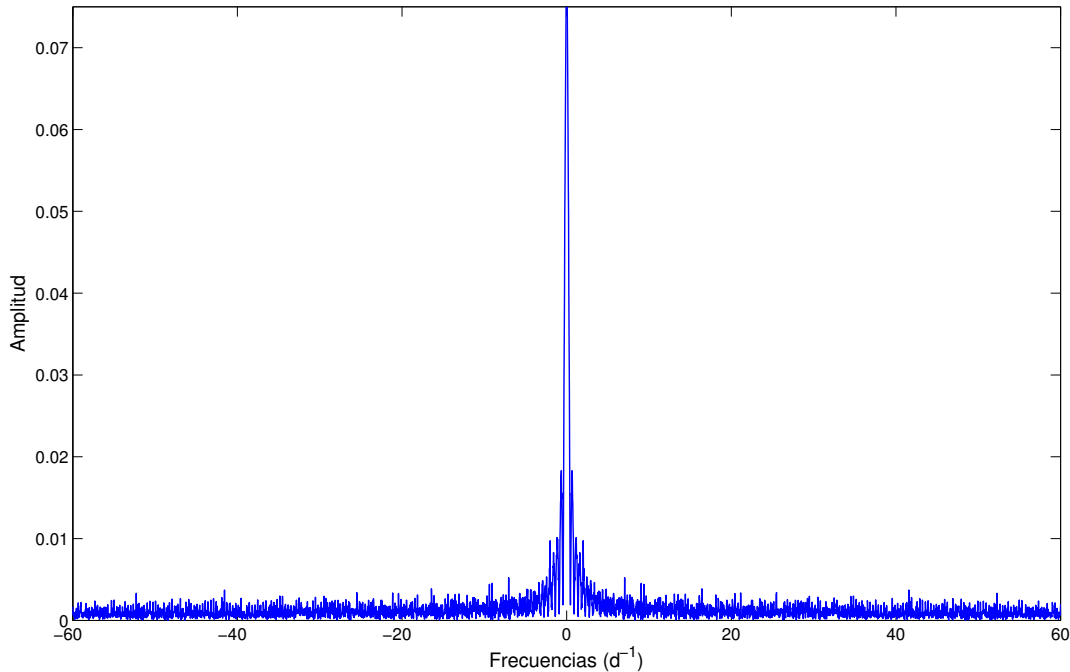


Figura 3.2: Ventana espectral típica en una curva de luz de *Kepler*. La serie consta de 41443 puntos con un muestreo de 60 s. aproximadamente haciendo un total de 30.79 días observados. Nótese que los alias aparecen con amplitudes muy bajas.

Por último, en la Fig. 3.3 observamos la ventana espectral de uno de los objetivos del campo de sismología de CoRoT. Profundizamos más en este análisis ya que es relevante para el estudio que presentamos en los próximos capítulos. Ésta es la ventana espectral típica de todos los objetivos observados durante ese run. En la figura se observan alias a 2 y 4 cd^{-1} y una serie de patrones centrados en torno a 13.97 cd^{-1} y múltiplos que constan de frecuencias secundarias separadas por 1 cd^{-1} y múltiplos. El alias de 2 cd^{-1} se debe al paso del satélite por la Anomalía Sur Atlántica (ASA) durante el cual el aumento del impacto de partículas cargadas introduce una perturbación sobre la CCD (Samadi et al., 2007). Ésto ocurre 2 veces durante un día sidéreo. A diferencia de *Kepler*, que sigue una órbita heliocéntrica, CoRoT sigue una órbita polar alrededor de la Tierra y por tanto se ve afectado por las variaciones orbitales y por los protones atrapados por el cinturón de Van Allen. Por otra parte, los alias de 13.97 cd^{-1} y múltiplos se deben a variaciones entre órbitas, ya que el satélite da 14 vueltas a la Tierra durante un día.

Para el caso de un muestreo uniforme podemos determinar con exactitud las propiedades de la ventana espectral dado el tiempo de muestreo y el tiempo de observación. A partir de la Eq. 3.3, teniendo en cuenta la linealidad del muestreo ($t_n = t_0 + \Delta t$) y eligiendo el origen t_0 de forma adecuada encontramos que la función ventana espectral es:

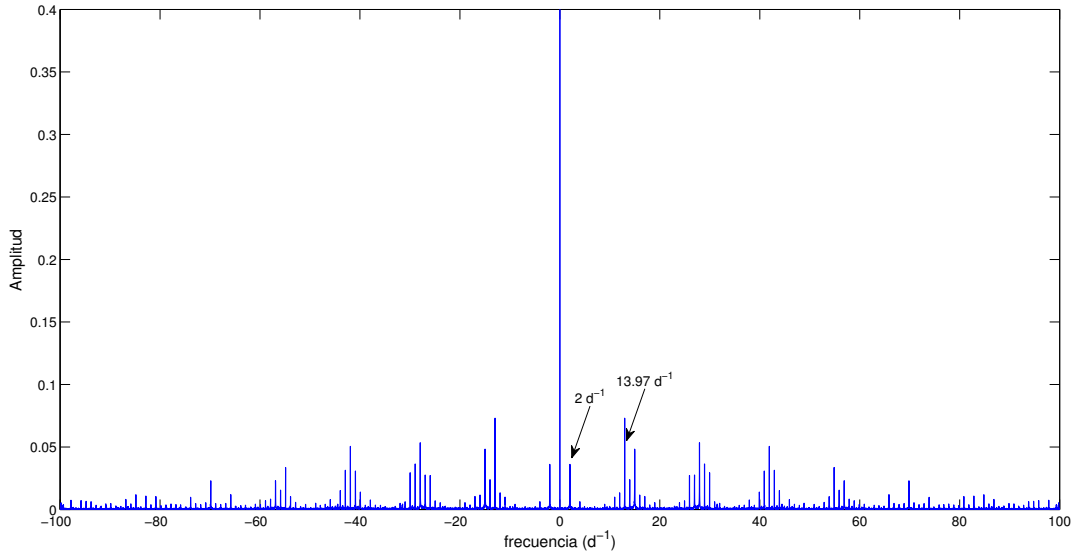


Figura 3.3: Ventana espectral de la observación de la estrella HD 50890 que fue observada durante el Initial Run (IRa01) de CoRoT durante 56.7 días. Nótese los alias a las frecuencias de 2 d^{-1} , a 13.97 d^{-1} y múltiplos de estas.

$$V_N(\nu) = \frac{\sin(\pi\nu N\Delta t)}{N \sin(\pi\nu\Delta t)} \quad (3.4)$$

A partir de esta ecuación podemos deducir que la ventana espectral cuando el muestreo es uniforme es una función simétrica y periódica con periodo Δt^{-1} . Combinando estas dos propiedades se puede ver que la forma de la función es la de un peine de Dirac sólo que en lugar de deltas de Dirac encontramos una serie de sincs espaciadas a intervalos de Δ^{-1} . Conforme el número de muestreos N tienden a infinito las sincs se convierten en funciones delta. En cambio, para series con un muestreo no uniforme las propiedades de simetría y periodicidad no se cumplen de modo que no podemos determinar con facilidad el efecto de la ventana espectral. Otro aspecto a destacar sobre la ventana espectral es la forma en que ésta modifica las fases de la señal original. En general la ventana espectral no es una función real y su fase introduce una desviación en la fase de la TFD sobre la transformada real $F(\nu)$. Esta desviación, que puede ser determinada para el caso de un muestreo uniforme, se complica para un muestreo no uniforme de manera que las fases pueden aparecer correladas dificultando los estudios en los que se requiere la determinación de las fases (p.ej. diferencias de fases entre curvas de luz de distintas longitudes de onda para la identificación modal [Garrido 2000](#)).

Un último problema sobre el muestreo que es necesario remarcar es el de la normalización. Para que la transformación tenga significado físico es necesario buscar una normalización adecuada. Podemos hablar de una normalización física cuando al tender N a infinito se obtiene un valor finito para la transformada ([Deeming, 1975](#)).

3. EFECTOS DEL MUESTREO EN LA DETECCIÓN DE SEÑAL: LA VENTANA ESPECTRAL

La normalización surge de un teorema matemático que puede considerarse una formulación alternativa del principio de conservación de la energía: el teorema de Parseval. Según este teorema:

$$k \sum_{i=1}^N |F_N(\nu)|^2 = \frac{1}{N} \sum_{n=1}^N f(t_n)^2 \quad (3.5)$$

donde k es el factor de normalización.

Cuando definimos anteriormente la TFD utilizamos la normalización de $1/N$ ya que esto permite que se cumpla la condición dada (siempre que tratemos con funciones absolutamente integrables). No obstante, para el caso de un muestreo no uniforme no es posible determinar las propiedades de normalización del mismo modo.

Algunas de las peculiaridades de la ventana espectral cuando el muestreo es no uniforme se han podido sortear a través de modificaciones sobre la expresión de la TFD definida en la Eq. 3.2. Así, el periodograma de Lomb-Scargle (Scargle, 1982), que es el más utilizado en la literatura de astrosismología de entre todas las versiones de periodogramas para la detección de frecuencias, permite recuperar las mismas propiedades estadísticas que se tiene con un muestreo uniforme - una χ^2 con 2 grados de libertad. Sin embargo, el periodograma de Lomb-Scargle está limitado a series en las que solamente haya ruido blanco y, además, no soluciona el efecto que tiene la presencia de huecos en el espectro, ya que es equivalente a la sustitución de éstos por ceros que generan la misma ventana espectral. Como consecuencia en el periodograma de Lomb-Scargle hay correlaciones entre las frecuencias que normalmente se ignoran.

Como ejemplo de correlaciones entre frecuencias introducidas por la ventana espectral supongamos que estamos observando un fenómeno aleatorio que describimos mediante la función $f(t)$ y que tiene una distribución de ruido blanco con varianza σ_0^2 y media cero. La observación viene dada por una ventana $v(t)$ de modo que la transformada de la función será:

$$F(\nu) = \int_{-\infty}^{+\infty} f(t) v(t) e^{i2\pi\nu t} dt$$

y por tanto la correlación media entre 2 frecuencias ν_1, ν_2 vendrá dada por

$$E \left[F(\nu_1) F^*(\nu_2) \right] = E \left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(t) f(t') v(t) v(t') e^{i2\pi\nu_1 t} e^{-i2\pi\nu_2 t'} dt dt' \right]$$

donde al introducir la transformada de Fourier inversa de $f(t)$ y $f(t')$ y utilizando las propiedades del ruido blanco se obtiene

$$E \left[F(\nu_1) F^*(\nu_2) \right] = 2\sigma^2 \int_{-\infty}^{+\infty} V(\nu_1 - \nu) V(\nu - \nu_2) d\nu$$

y finalmente, usando las propiedades de convolución de la transformada de Fourier inversa

$$E\left[F(\nu_1)F^*(\nu_2)\right] = 2\sigma^2 \int_{-\infty}^{+\infty} v^2(t)e^{i2\pi(\nu_1-\nu_2)t} d\nu \quad (3.6)$$

La integral es la transformada de Fourier del cuadrado de la función ventana $v(t)$. Como ejemplo más simple, para una función continua con una ventana de observación rectangular la correlación media entre frecuencias es una sinc.

La expresión dada en la Eq. 3.6 es válida sólo para un proceso estocástico tipo ruido blanco pero, cuando $E[F(\nu)^2]$ varía lentamente (más lentamente que $1/T$ por ejemplo), entonces esta expresión nos da también una idea aproximada de lo que está ocurriendo. Es decir, la ventana con muestreo no uniforme puede introducir fuertes correlaciones entre las componentes a distintas frecuencias, tanto en sus amplitudes como en sus fases, causando así múltiples complicaciones en el análisis como, por ejemplo, el ajuste de los parámetros correspondientes a cada línea (anchura, centroide, etc.) que permite comparar los modelos teóricos y las observaciones.

Otra de las ventajas que tiene el muestreo uniforme de la función es que nos permite hallar su función de autocorrelación. Cuando analizamos una función estocástica o una función combinada que contiene una componente estocástica, podemos hallar su espectro de potencias mediante una TFD de la función de autocorrelación pero solamente si el muestreo es uniforme ya que de lo contrario no hay forma de hallar la autocorrelación.

En la tabla 3.1 mostramos un compendio de todas las propiedades mencionadas por las que es, como mínimo, útil disponer de un muestreo uniforme de la función previamente al análisis. Por rigor científico es de notar que el muestreo no uniforme tiene propiedades que lo hacen deseable en determinadas situaciones. Si el observador puede elegir los tiempos de muestreo es posible configurar la ventana espectral de forma que esta afecte lo mínimo posible al análisis de frecuencias. Así, por ejemplo, se pueden reducir los alias producidos por altas frecuencias en el rango de bajas frecuencias controlando el muestreo. No obstante, en la práctica es poco probable que se pueda elegir los tiempos de muestreo y por otro lado, cuando se tiene un muestreo no uniforme el límite de Nyquist no se elimina por completo (Bretthorst, 2000). De hecho, el muestreo no uniforme solamente es completamente eficiente eliminando los alias debido a frecuencias más allá de Nyquist para el caso de un muestreo completamente aleatorio y $N \rightarrow \infty$.

De las propiedades que aparecen en la tabla 3.1 en especial la última es la que queremos remarcar aquí ya que tendrá especial relevancia para el estudio que hacemos en

3. EFECTOS DEL MUESTREO EN LA DETECCIÓN DE SEÑAL: LA VENTANA ESPECTRAL

la última parte de la memoria. De hecho, es imprescindible obtener un muestreo uniforme para poder reconstruir la señal de manera natural conservando sus propiedades intrínsecas y así poder caracterizar la función de manera local. De este modo, además de las propiedades mencionadas, el estudio que realizamos en la Parte IV de la memoria, que motiva el título de esta tesis, justifica la necesidad de interpolar la serie en los huecos para obtener un muestreo uniforme.

Hasta aquí se ha definido la ventana espectral y se ha mostrado la patología que ésta recoge. Sin embargo, la patología del espectro de frecuencias de una serie temporal no se restringe sólo a la que se origina por el modo en que se distribuyen los datos, interviene también el tipo de señal que se analiza y las hipótesis que subyacen al propio análisis. La filosofía que se sigue habitualmente cuando se lleva a cabo un análisis de Fourier es interpretar los datos de acuerdo a un modelo matemático preestablecido que representa el modelo físico del fenómeno estudiado. En el caso que nos ocupa, el de las estrellas pulsantes multiperiodicas, se parte de un modelo determinista basado en la composición de un ruido blanco con distribución gaussiana y una (o varias) de estas 3 formas funcionales:

- Función periódica de tipo no-sinusoidal
- Función sinusoidal
- Oscilador armónico excitado estocásticamente

Como si fuera el modelo teórico el que determinase la observación y no viceversa, estas formas funcionales, que se derivan de modelos físicos para cada clase de estrella pulsante, determinan la forma de analizar las correspondientes curvas de luz observadas. Así pues, curiosamente, hasta ahora no se han hecho análisis partiendo de diferentes formas funcionales. Pero es esencial tener en cuenta las limitaciones que esto impone al análisis. En general, un modelo determinista se comporta de forma distinta a uno no-determinista bajo un análisis de Fourier. Dentro de los modelos deterministas las funciones periódicas, se comportan de forma distinta a las no-periódicas. Es más, la propia elección de una transformada de Fourier como herramienta de análisis ya introduce una hipótesis fuerte que limita el modelo, esto es que la función sea absolutamente integrable, ya que esta es la condición para que la función se pueda representar mediante una transformada de Fourier en el espacio de frecuencias.

Dadas las limitaciones comentadas del modelo matemático, el criterio de corrección científica nos obliga a considerar un método para obtener la representación más fidedigna de la función en el espacio de frecuencias. Esto implica evitar las interferencias debidas a la ventana espectral sin introducir ninguna variación en la serie que

no sea propia de la señal. Con esto en mente, realizamos un relleno de los huecos para obtener un muestreo uniforme que, si bien también conlleva efectos de interferencia en el espectro de frecuencias, estas están perfectamente determinadas. Se han propuesto múltiples procedimientos para llevar a cabo el relleno de huecos pero esta metodología no está exenta de complicaciones e inconvenientes como veremos en el próximo capítulo.

3. EFECTOS DEL MUESTREO EN LA DETECCIÓN DE SEÑAL: LA VENTANA ESPECTRAL

TABLA 1.	La ventana espectral con muestreo uniforme
1)	Teor. de Shannon - reconstrucción de la señal de forma exacta.
2)	Simplifica el análisis: la ventana espectral es una serie de sincs que tienden a un peine de Dirac cuando N tiende a ∞
3)	Elimina las contribuciones espurias irregulares que puede haber lejos del pico principal
4)	Es posible describir las propiedades de la ventana espectral (i.e. simetría y periodicidad)
5)	La desviación sobre las fases que produce la ventana espectral puede ser corregida fácilmente a diferencia de lo que ocurre con el muestreo no uniforme.
6)	Normalización - este problema es solamente trivial para un muestreo uniforme
7)	La ventana no introduce correlaciones entre frecuencias
8)	Función de autocorrelación - esta sólo se puede determinar con un muestreo uniforme
9)	Caracterización de las propiedades locales de la función - esto sólo se puede hacer cuando la función se puede reconstruir de manera exacta, esto es con un muestreo uniforme adecuado.

Tabla 3.1: Propiedades de la ventana espectral de una serie con muestreo uniforme respecto a una serie con muestreo no uniforme

Capítulo 4

Técnicas de relleno de huecos

*“Sometimes it is necessary to extrapolate” Dr Brown rejoined testily,
“and even speculate, based on minimal amount of data. You know yourself
that new science is based on maximum likelihood rather than certainty”*

Rama II - Arthur C. Clarke

4.1 Métodos de Interpolación

En este apartado daremos una perspectiva general de los principales métodos de interpolación para relleno de huecos. Para desarrollar el argumento que nos ha llevado a buscar un nuevo método de interpolación -la necesidad de reconstruir la señal sin usar hipótesis de base ni introducir artefactos- discutiremos más a fondo 3 de estos métodos: interpolación lineal, inpainting, y el método de Fahlman y Ulrych.

Los métodos de relleno de huecos coinciden todos en su finalidad, que es reemplazar los huecos por predicciones de los segmentos de datos correspondientes de forma que preserven el contenido espectral de la serie. Además, todos ellos parten de que todos los segmentos de datos intercalados entre huecos todos tienen un muestreo uniforme. Además, éste suele ser el caso en las observaciones astronómicas, tanto desde satélite como desde tierra. Aunque las series temporales con huecos se pueden considerar un caso particular de muestreo no-uniforme distinguiremos aquí entre series no-uniformemente espaciadas y series con huecos definiendo a las primeras como aquellas cuyo muestreo $t_{i+1} - t_i = \Delta_i$ varía de forma irregular para cada punto. Teniendo en cuenta lo dicho, en adelante ignoraremos este caso y consideraremos sólo el caso de series con huecos para el estudio de los métodos de interpolación.

4. TÉCNICAS DE RELLENO DE HUECOS

4.1.1 Interpolación lineal

Es el tipo de corrección que se ha aplicado a los datos de nivel 2 de CoRoT (Samadi et al., 2007). Como se mencionó en el capítulo anterior, durante el paso por la Anomalía Suratlántica (AS) los datos obtenidos se eliminan con lo que se pierden en la mayoría de los casos en torno a un 10 % de los puntos. Los huecos resultantes se rellenan posteriormente durante la reducción de datos mediante interpolación lineal.

Si bien en algunos trabajos se ha optado por hacer un análisis de frecuencias basado en la estimación del espectro de potencias de la serie con huecos sin rellenar (e.g. Mantegazza et al. 2012; Rauer et al. 2009; García Hernández et al. 2013; en su análisis de las estrellas HD 50870, CoRoT-5b, HD 174966, respectivamente), el método de interpolación lineal es el procedimiento aplicado en la mayoría de los casos en que se trabaja con datos de CoRoT (e.g. Appourchaux et al. 2008; Benomar et al. 2009, Zwintz et al. 2011; Deheuvels et al. 2010; García et al. 2009; en los análisis de las estrellas HD 49933, V 588 Mon y V 589 Mon, HD 49385, HD 181906 respect.). No obstante, este método es poco coherente con la información de la señal y en ciertas situaciones puede resultar problemático, especialmente cuando hay componentes armónicas de bajas frecuencias como en HD 175726 (Mosser et al., 2009). En este caso se ha usado un algoritmo basado en *wavelets* (Holschneider et al., 1990) para la reconstrucción de los huecos. Sin embargo, los resultados (cf. fig. 3 en Mosser et al. 2009) muestran que sólo las bajas frecuencias se recuperan de manera adecuada y, en la práctica, este algoritmo proporciona sólo una forma más sofisticada de hacer una interpolación polinómica que no es fiel a la información contenida en la señal. Otros autores han optado por interpolar parábolas (Ballot et al., 2011), que es también otro tipo de interpolación polinómica y sólo corrige bien las frecuencias muy bajas.

En algunos casos se ha evaluado el coste de introducir una interpolación lineal (e.g. Gutiérrez-Soto et al. 2009; Neiner et al. 2009; Appourchaux et al. 2008; en los análisis de las estrellas HD 175869, HD 181231, HD 49933 respect.) y en general, en éstos se concluye que el impacto es mínimo para las frecuencias bajas y sólo una pequeña reducción de la amplitud de las altas frecuencias en torno al 10%. Sin embargo, si bien ésto es cierto en lo que respecta a las amplitudes de las componentes afectadas de la señal, no se tiene en cuenta un efecto de tipo estructural. Es decir, no se ha considerado que la interpolación, además de afectar a las amplitudes, también repercute sobre la forma de la ventana espectral y las frecuencias espurias que aparecen a causa de ésta. De hecho, el relleno de huecos más simple, que consiste en generar un vector de tiempos t_i e insertar ceros en los valores en que no hay medidas para obtener un muestreo uniforme, da lugar a una serie cuyo espectro es equivalente al que resulta utilizando el periodograma Lomb-Scargle para analizar la serie con muestreo

no-uniforme. Es decir, la reducción del efecto de la ventana espectral es nula mediante este método que puede considerarse un caso especial de interpolación lineal.

Por otra parte, se puede comprobar que en la mayoría de los casos, la interpolación lineal supone una mejora poco significativa sobre el relleno con zeros y, a diferencia de ese caso, tiene la complicación añadida de que no es trivial hallar qué forma tiene la función de la ventana espectral que convoluciona la señal.

Puesto que nos interesan los datos de la misión CoRoT y este es el método de interpolación más utilizado en la literatura sobre estos datos, es el que estudiamos más a fondo y que usamos como referencia para comparar con el método de interpolación ARMA en la § 5.1. Los efectos que produce este tipo de interpolación sobre el espectro serán analizados con más detalle posteriormente.

4.1.2 Inpainting

Este método se ha implementado en el pipeline para el análisis de estrellas tipo solar (Mathur et al., 2010), y es uno de los métodos más utilizado recientemente en la literatura de CoRoT (Uytterhoeven et al. 2011, Mathur et al. 2010, Mathur et al. 2013), de ahí que le dediquemos una sección aparte.

El método de Inpainting forma parte de un tipo de técnicas conocidas como Análisis de Componentes Morfológicas (en inglés MCA; Elad et al. 2005) introducido para el relleno de agujeros en imágenes bidimensionales y se ha utilizado además de en astrosismología, en el estudio de la deformación del espacio por lentes gravitacionales débiles (Pires et al., 2009), y en el estudio de las imágenes de la radiación cósmica de microondas (Starck et al., 2013). Este método se basa en la existencia de un diccionario Φ en el que la representación de los datos completos viene dada por una matriz en la que la mayoría de los valores son cero o cercanos a cero (matriz dispersa), y la representación de los datos incompletos viene dada por una matriz no dispersa, es decir, cuyos valores son distintos de cero. Se trata, por tanto, de minimizar

$$\min_X \|\Phi^T X\|_1 \quad \text{sujeto a} \quad \sum_i (Y - MX)^2 \leq \sigma$$

donde X es la serie completa, Y es la serie observada, M es la ventana ($M_i = 1$ para los datos válidos, y $M_i = 0$ para los huecos) y σ es la desviación estándar. Este método consiste en recuperar X conocidos Y , M . Para ello, en el caso que nos ocupa, se utiliza un diccionario basado en Transformadas de Coseno Discretas (TCD), que son transformadas relacionadas con la de transformada de Fourier. Previamente se usa una descomposición wavelet para determinar el rango de frecuencias que se puede interpolar en cada hueco variando el tamaño de la TCD local en relación con las

4. TÉCNICAS DE RELLENO DE HUECOS

distintas escalas de los huecos presentes en la serie, esto es lo que se conoce como una Transformada de Coseno Discreta Multiescala.

En definitiva, el método de Inpainting interpola la información perdida usando una información a priori, que en este caso se trata de que la matriz que representa la señal sea una matriz dispersa con el diccionario Φ basado en transformadas TCD. Así pues, este método describe los datos en los huecos gracias a una extrapolación basada en el contenido en frecuencias derivado de la TCD.

La información a priori consiste en la hipótesis de partida de que la serie se puede ajustar con determinadas funciones de forma cerrada, pero además lleva otras presunciones sobre los datos como la analiticidad de la función que describe a los datos, o dicho de forma aproximada, que la función sea suave. Por otro lado, los resultados del artículo, [Sato et al. 2010](#), en el que se introduce éste método para la corrección de series en astrosismología, muestran que este método no reproduce de forma sensible las altas frecuencias.

4.1.3 El método de Fahlman y Ulrych (MFU)

Se publicó en 1982 por [Fahlman and Ulrych \(1982\)](#) y se basa en el método de estimación espectral de Máxima Entropía (ME), introducido por J. P. Burg en su tesis doctoral 7 años antes ([Burg, 1975](#)). La idea detrás del método de análisis espectral ME es extrapolar la función de autocorrelación de tal modo que la entropía de la función densidad de probabilidad correspondiente sea maximizada en cada paso de la extrapolación. Esto se traduce en un principio de mínimo compromiso con la información no disponible, lo que se satisface no haciendo hipótesis alguna sobre los segmentos de la serie que no se han observado. Así pues, este principio resulta ser el más solidario con el criterio de corrección científica que hemos definido en la parte [2.3](#). Es por ello que el método que introduciremos para el relleno de huecos presenta una cierta analogía con éste. El método ME está intrínsecamente relacionado con los modelos autoregresivos y es esta relación en la que Fahlman y Ulrych se basan para el relleno de huecos de series temporales.

El algoritmo consiste en ajustar un modelo autoregresivo (ver def. en la siguiente sección) de forma global a los datos disponibles y realizar una predicción de este modelo sobre los huecos. Cada hueco se rellena de forma independiente a partir de los datos que lo preceden y que le siguen. El procedimiento es iterativo de modo que tras rellenar todos los huecos se ajusta un nuevo modelo con los datos originales más la predicción y se repite el proceso iteración tras iteración hasta que el error de predicción alcance un valor mínimo.

Este método se ha usado en diversas ocasiones de forma bastante exitosa (i.e. [Papparo and Kovacs 1984](#), [Salakhutdinova 1998](#), [Korzennik et al. 2013](#)). En [Schwarzenberg-Czerny \(1987\)](#) se prueba la superioridad del método de estimación espectral ME sobre las técnicas de Fourier. En [Carbonell et al. \(1992\)](#) este método se compara con distintas técnicas para la estimación espectral de series con huecos y se encuentra que es el método más eficiente recuperando el espectro original de la serie sin huecos salvo en el caso en que estos estén distribuidos de manera irregular. En este caso el algoritmo CLEAN introducido por [Högbon \(1974\)](#) funciona igual o mejor que el MFU. En el caso que nos ocupa, fundamentalmente los datos de CoRoT, los huecos se repiten de manera bastante regular de modo que es de esperar que el MFU funcione mejor.

El MFU no es sólo para el relleno de huecos sino que proporciona también una estimación espectral distinta a la de los métodos de Fourier y con mejor resolución espectral. Como estimador espectral tiene también la ventaja de que no requiere hacer ningún pre-whitening y que se puede aplicar incluso en los casos en los que la señal sea cuasiperiódica o no sea armónica.

Uno de los supuestos de este método es que los datos observados describen un proceso estacionario con un ruido blanco de distribución gaussiana. La estacionaridad del proceso está implícita en el hecho de que cada segmento de datos tiene el mismo contenido espectral y lo mismo para segmentos no observados, es decir, los huecos. Así pues, se utiliza un único proceso de un orden determinado para modelar todos los datos disponibles. El orden máximo está restringido por el tamaño del segmento de datos más corto. Esto supone una limitación bastante seria ya que de haber segmentos de corta duración entre huecos el orden máximo podría ser insuficiente para modelar toda la serie de forma adecuada. De hecho, en general se necesitan modelos con un orden bastante alto para representar los procesos con un número considerable de componentes armónicas. Por otro lado, la presencia de ruido limita la precisión en las amplitudes de las componentes armónicas pero, como hemos dicho, se presupone ruido blanco además, y de no ser así esto podría afectar en gran medida a otros aspectos del análisis.

Otro problema que surge en la práctica es que no existe un criterio objetivo para hallar la bondad del ajuste de un modelo AR, y esto afecta a la elección del orden del modelo y al número de iteraciones necesarias para finalizar el procedimiento, de modo que no se pueden fijar de manera rigurosa.

4.1.4 Otros métodos

Los métodos anteriores funcionan bien para series con una cobertura de datos por encima del 50% del periodo de observación y una razón señal-ruido suficiente. En la literatura se encuentran otros métodos que funcionan mejor en casos extremos en

4. TÉCNICAS DE RELLENO DE HUECOS

que los que la cobertura es menor, sin embargo, todos ellos funcionan en situaciones particulares (véase e.g. el método de música repetitiva de [Fossat et al. 1999](#) para los datos de SoHO).

En [Brown and Christensen-Dalsgaard \(1990\)](#) se afirma que el MFU no es apropiado para el espectro de las oscilaciones solares (tampoco lo sería para estrellas tipo-solar) dado que es un espectro muy complejo y se requieren modelos de orden muy alto para modelar los procesos armónicos involucrados. Para solventar los problemas que surgen de ahí introducen un MFU modificado en el que se hace uso de un filtro paso-banda que reduce la complejidad de la serie temporal, i.e. número de frecuencias de la señal. Esta versión modificada se ha venido utilizando desde entonces ([Anderson, 1993](#); [Toner et al., 1999](#)), sin embargo, los problemas computacionales señalados en [Brown and Christensen-Dalsgaard \(1990\)](#) no son hoy en día un gran obstáculo. Sí que podría serlo el hecho de no tener suficientes puntos para ajustar el modelo AR si el orden óptimo fuese demasiado alto. No obstante, en dicho trabajo se sobreestima el número de coeficientes necesarios pues se afirma que para ajustar una componente sinusoidal de la curva de luz se necesitan media docena de coeficientes, y como se verá más adelante basta con un modelo AR(2). Por otro lado, el número de coeficientes puede aumentar sin tener que hacer uso de más puntos si en lugar de usar modelos AR se usan modelos ARMA como es nuestro caso. Esa es, de hecho, una de las principales ventajas del método que se expondrá a continuación. Puede demostrarse, además, que la idea de que sea necesario un gran número de coeficientes para ajustar la curva de luz de las oscilaciones solares dado el número de frecuencias detectadas en el análisis de Fourier, es una suposición errónea.

El método de interpolación ARMA que introduciremos solventa algunas de las complicaciones de los métodos que hemos comentado anteriormente pero éste se ha diseñado para trabajar con los datos de los satélites CoRoT y *Kepler*, cuyas series no tienen las mismas interrupciones que desde tierra. De modo que el método ARMA no está orientado a rellenar huecos grandes ni observaciones con una escasa cobertura de datos, sino que más bien este método tiene como objetivo reproducir la señal en los huecos sin hacer uso de funciones de forma cerrada. De este modo se evitan posibles incoherencias originadas por las preconcepciones sobre la señal.

El método de interpolación ARMA cumple las siguientes características:

- No requiere estacionaridad ya que el análisis es local.
- Es más eficiente, ya que los modelos ARMA convergen más rápidamente.
- Si el orden AR es n el número de parámetros puede aumentar hasta $2n$ sin necesidad de hacer uso de un mayor número de puntos.

- Puede modelar una serie con ruido coloreado además del ruido blanco.
- Puede incluso, modelar una serie que provenga de una función no diferenciable.
- No requiere ninguna información a priori sobre el tipo de función que modela.

4.2 Modelos autoregresivos

Los modelos autoregresivos son una representación paramétrica de las serie de datos que da cuenta de la autocorrelación de la serie en forma de ley recursiva. Así, siendo una representación tan general puede ser utilizada en la mayoría de los casos sin restricciones significativas. El modelo autoregresivo más sencillo fue introducido como representación discreta de una señal armónica perturbada (Yule, 1927). Así, la ecuación diferencial del oscilador armónico amortiguado excitado estocásticamente

$$\ddot{x}(t) = -\eta\dot{x}(t) + \omega^2x(t) + \epsilon(t)$$

donde η es el amortiguamiento, ϵ es el término de excitación y ω es la frecuencia de la oscilación, puede ser discretizado mediante un proceso AR de segundo orden tal como este:

$$x_n = a_1x_{n-1} + a_2x_{n-2} + \epsilon_n \quad (4.1)$$

En este caso el problema se resuelve fácilmente dando lugar a estas expresiones para los coeficientes:

$$a_1 = 2e^{-1/\tau} \cos(\omega) \quad a_2 = -e^{-2/\tau}$$

donde τ es el tiempo de amortiguamiento de la oscilación (Roth and Zhugzhda, 2010). Para una serie temporal originada en una observación de una estrella pulsante multiperódica como el Sol con M componentes armónicas, se podría hallar una representación general en forma de un proceso AR simplemente sumando las M ecuaciones de los procesos autoregresivos individuales. Así pues, un proceso AR puede modelar oscilaciones estocásticas, es decir, del tipo que encontramos en el Sol, pero también se pueden modelar otros fenómenos. De hecho, los procesos AR son modelos muy generales que pueden usarse para modelar casi cualquier tipo de variación. Sin embargo, el uso y la comprensión de los procesos autoregresivos a menudo se ha limitado a otros dominios de la ciencia distintos a la astrofísica por diversas preconcepciones equivocadas sobre esta técnica. Una de las preconcepciones sobre este tipo de modelos es que sólo pueden utilizarse para modelar procesos estocásticos pero esa idea debe ser abandonada en base al modelo general de serie temporal que propone el Teor. de

4. TÉCNICAS DE RELLENO DE HUECOS

descomposición de [Wold \(1938\)](#). Este teorema dice que cualquier proceso estocástico estacionario permite la descomposición:

$$x_n = u_n + v_n$$

donde:

- v_n es un proceso determinista
- u_n es un proceso no-determinista con una distribución espectral absolutamente continua

u_n y v_n son estacionarios y no están correlados entre ellos. Esta descomposición coincide con el modelo base que se utiliza para cualquier análisis de una observación, pues en ésta siempre hay una componente determinista, que es la que se trata de analizar, y una componente aleatoria o puramente estocástica debido a ruido o errores en la medida. De modo que, puesto que en ningún caso se obtiene una serie puramente determinista, según este teorema siempre podemos considerar que las series observadas son estocásticas, independientemente de que estas contengan una componente determinista o no. Algunos autores en astrosismología consideran, aunque con una notable desviación, que solamente son estocásticas las series que proceden de oscilaciones excitadas de manera estocástica, como es caso de las estrellas pulsantes de tipo solar. En este sentido, los modelos AR, que se utilizan para modelar fenómenos estocásticos, se pueden utilizar también para fenómenos con una componente determinista.

Por otro lado, también se ha considerado a menudo que la definición paramétrica del modelo supone una restricción al tipo de serie representable, y se desconoce a priori si los datos cumplen las condiciones para ser modelados de esa forma ([Stoica et al., 2000](#)). No obstante, como hemos visto, proporcionan una representación muy general y su uso solamente está limitado por la estacionaridad de la señal. En el caso de oscilaciones con una fuerte modulación temporal, por ejemplo, los coeficientes del modelo autoregresivo variarían con el tiempo y sería necesario hacer uso de modelos más complejos.

Los procesos AR se pueden generalizar, como veremos. La componente no-determinista u_n en la ecuación anterior se puede representar como un proceso de media móvil (MA, del inglés Moving Average):

$$u_n = \sum_{k=0}^q b_k e_{n-k}, \quad b_0 = 1 \quad (4.2)$$

donde e_n son términos de una variable aleatoria no correlada, con las propiedades $E[e_n] = 0, E[e_n^2]$, es decir ruido blanco. En este caso hemos utilizado un modelo MA de orden q para representar el proceso u_n . La operación que aparece en la ecuación anterior se puede describir mediante un operador B de retroceso, que cambia e_n en e_{n-1} . Invertiendo ese operador se obtiene una representación a través de un proceso AR que es equivalente al modelo MA. Es decir, un modelo AR es siempre equivalente a un modelo MA siempre que sea invertible. Pero incluso en los casos en que el proceso en cuestión no sea invertible se puede encontrar un modelo equivalente que sí lo sea (Ulrych and Clayton, 1976). La ventaja de escribir un proceso MA en forma de proceso AR es que es mucho más sencillo determinar los parámetros del modelo AR.

De la misma forma que un proceso MA se puede invertir y transformar en un AR, también es posible hallar modelos mixtos ARMA, que tienen componentes AR y componentes MA tal como

$$x_n = \sum_{k=1}^p a_k x_{n-k} + \sum_{k=1}^q b_k e_{n-k} \quad (4.3)$$

Un modelo ARMA es similar a un modelo AR pero asume un ruido. A diferencia del modelo AR, la componente MA del modelo ARMA hace posible modelar los ceros que aparecen debido al ruido observado. Ya que todas las series que encontramos tienen una componente de ruido en mayor o menor amplitud, el modelo ARMA da cierta ventaja pese a que añade cierta complejidad al modelado. Particularmente se ha encontrado que las estimaciones del espectro que pueden obtenerse usando modelos ARMA para representar funciones sinusoidales con ruido aditivo son superiores a las que dan los modelos AR (Ulrych and Ooe, 1979).

Uno de los motivos centrales de la relevancia de los modelos autoregresivos es su correspondencia con el método de estimación espectral de Máxima Entropía (ME) de Burg (1975). Esto supone que se cumple el principio del mínimo compromiso con la información no disponible.

Los modelos autoregresivos resultan muy útiles para hacer predicción pero la predicción de largo alcance pierde coherencia y su amplitud decae rápidamente. Por tanto, sus ventajas son aprovechables para el relleno de huecos de tamaño corto y mediano. Se han propuesto variaciones para subsanar este problema (p.ej. Esquef et al. 2003; Esquef and Biscaíno 2006) pero éstas no cumplen tan bien el principio de mínimo compromiso con la información no disponible y en nuestro caso no es necesario hacer uso de ellos ya que los huecos de CoRoT y Kepler no tienen un tamaño muy grande.

4. TÉCNICAS DE RELLENO DE HUECOS

En general AR, MA, y ARMA son modelos autoregresivos que pertenecen a una clase más general conocida como ARIMA donde la I se refiere a un tercer término integral que permite ajustar cierta no-estacionaridad en las series. Aquí no tenemos en cuenta los modelos ARIMA puesto que en las observaciones de CoRoT y *Kepler* la no-estacionaridad no es significativa para el relleno de huecos, ya que éstos son de una escala mucho menor que la de las variaciones estacionales.

4.3 El Método de Interpolación ARMA

El procedimiento para el relleno de huecos que usamos aquí (Pascual-Granado et al., 2012) es similar al de Fahlman y Ulrych pero usamos modelos ARMA en lugar de AR, y en lugar de utilizar un único modelo global ajustado mediante todos los segmentos de datos disponibles, nosotros hallamos modelos locales a partir de los segmentos de datos en torno a cada hueco. Esto permite reproducir las variaciones pequeñas de la serie, admite cambios estacionales en la señal, y mantiene mejor la coherencia temporal, con lo cual solventamos de una vez todos los problemas que puede achacarse a los modelos autoregresivos.

La aproximación local se ha usado anteriormente en Roques et al. (2000), sin embargo, allí usaron modelos AR, que pueden requerir un gran número de términos y el modelo que se usa entonces es solamente una aproximación. Además, la forma de obtener el orden del modelo que se usa Roques es heurística, es decir, se fija el orden entre $1/3$ y $1/2$ del tamaño medio de los segmentos de datos contiguos a los huecos. Este método no utiliza la información presente en la serie y nosotros desarrollamos un criterio para la elección del orden basado en propiedades estadísticas de los datos a estudiar.

El código para la implementación del Método de Interpolación ARMA (MIARMA) se ha programado en el Entorno de Desarrollo Integrado de la herramienta de software matemático MATLAB.

Para la descripción del método seguiremos aquí una notación similar a la que se usa en Fahlman and Ulrych (1982). Si la serie observada es x_n con $n \in (1, N)$, decimos que hay S segmentos de datos con muestreo uniforme con el índice $n \in (M_s + 1, N_s)$ siendo $s \in (1, S)$, y $M_1 = 0$, $N_S = N$. Por otro lado, los $S - 1$ huecos tienen como índices $n \in (N_s + 1, M_{s+1})$ con $s \in (1, S - 1)$.

4.3.1 Criterio para la elección del orden

Para ajustar modelos a los segmentos de datos el primer paso es la elección del orden, es decir, el número de términos autoregresivos (p) y el número de términos de media móvil (q), dando lugar a un modelo ARMA(p,q) tal y como se muestra en la

Eq. 4.3. La determinación de los órdenes es crucial para que la representación sea correcta puesto que del orden depende si el modelo puede ajustar todos los detalles de la serie o no. Valores demasiado pequeños del orden dan modelos con insuficiente resolución espectral, mientras que valores demasiado altos dan lugar a situaciones de inestabilidad y a la aparición de detalles espúreos.

El Criterio de Información de Akaike (1974) da una medida de la calidad de un modelo en relación a un segmento de datos determinado. Esto permite seleccionar el orden más adecuado para el segmento de datos modelado. El criterio se basa en maximizar la función de verosimilitud de los parámetros, o lo que es lo mismo minimizar el error cuadrático medio de la predicción (Ulrych and Ooe, 1979). El parámetro AIC dado se define como

$$AIC = N \cdot \log V + 2 \cdot d$$

donde V es la suma de los cuadrados de los residuos de un modelo dado, d es el número de parámetros del modelo (para un proceso ARMA sería $d = p + q$, siendo p el orden AR y q el orden MA) y N el número de datos modelados. Para obtener V se debe realizar el ajuste del modelo ARMA(p, q) cada vez mediante el procedimiento que se describirá en la próxima sección.

De acuerdo a la teoría de Akaike el modelo óptimo es el que da el menor valor del coeficiente AIC. Por tanto, el criterio consiste en computar el coeficiente AIC para un conjunto de modelos con distintos ordenes p y q , y escoger aquellos para los cuales se minimice este valor.

En Ulrych and Clayton (1976) se afirma que el modelo ARMA(p, p) es la representación más parsimoniosa (i.e. que requiere el mínimo número de coeficientes - Scargle) de un proceso armónico con ruido blanco gaussiano, de modo que para modelar este tipo de procesos tiene sentido seleccionar modelos ARMA con el mismo orden para la parte AR que para la parte MA. No obstante, puesto que en nuestro caso queremos mantener la generalidad y no introducir ninguna hipótesis ad hoc sobre el tipo de proceso que se observa, iteramos el algoritmo sobre todos las combinaciones p y q posibles dentro de unos límites predefinidos para hallar el valor AIC óptimo. Para hacer una estimación del rango sobre el que se deben iterar los órdenes p y q en (Ulrych and Ooe, 1979; Kovács, 1983) se propone una regla empírica que restringe el orden p del modelo óptimo al rango

$$N/3 - 1 \leq p \leq N/2 - 1$$

Por otro lado, el orden q en el caso de modelos ARMA se evaluaría en el rango $(0, p) \equiv (0, N/2 - 1)$. De este modo, para un segmento de datos de tamaño $N = 20$ el rango de órdenes (p, q) obtenido sería (redondeando) desde el $(6, 0)$ hasta el $(9, 9)$,

4. TÉCNICAS DE RELLENO DE HUECOS

es decir, un total de 40 modelos ARMA serían evaluados. Esta regla proporciona un método práctico para elegir el conjunto de órdenes (p, q) sobre los que evaluar el modelo ARMA, sin embargo, para series de datos con N muy grande el coste computacional es demasiado alto y, dada la falta de rigor matemático del método, es preferible utilizar alguna información a priori para elegir el rango más adecuado. En nuestro caso, dado que el tamaño medio de los huecos que quedan al eliminar los datos inválidos de las curvas de luz de CoRoT es de ~ 20 puntos exploramos todas las combinaciones posibles de p y q hasta el orden $(60, 60)$. No obstante, un rango diferente se puede introducir como entrada de la subrutina de estimación del orden ([armaord.m](#)).

Existen muchos otros criterios de selección del orden en la literatura como el Criterio de Akaike corregido (AICc), el Criterio de Información Bayesiana (BIC), el Error de Predicción Final (FPE), o el Criterio de Hannan-Quinn (HQ). Al igual que con el AIC, el parámetro a evaluar correspondiente a cada uno de ellos se compone de 2 términos: el primero ($N \log V$ en el caso de AIC o AICc) caracteriza la tasa de entropía o error de predicción del modelo; el segundo caracteriza el número de parámetros libres usados en el modelo. Minimizando los 2 términos se asegura que el modelo sea parsimonioso y por tanto no se produce *overfitting*, y se garantiza que el ajuste de los datos sea óptimo. Lo que diferencia a cada uno de estos modelos es el segundo término, en el que se pesa de diferente manera el número de parámetros d . El Criterio de Akaike introduce una penalización sobre modelos con d muy grande que es menor que la de los demás criterios. Esto no supone ningún problema en nuestro caso ya que nos restringimos a rangos de d que no son muy grandes. Por otra parte, el Criterio de Akaike es un criterio que se deriva de consideraciones puramente físicas, de hecho, Akaike denominó originalmente a esta aproximación el Principio de Máxima Entropía ya que es equivalente a maximizar la entropía de un sistema termodinámico. Es por ello que preferimos utilizar este criterio en lugar de los otros.

Finalmente, el uso de este criterio garantiza encontrar el orden óptimo para el modelo ARMA de una serie de datos finita con una relación señal-ruido determinada. Estas propiedades de la serie configuran el modelo ARMA óptimo, no obstante, la representación ARMA del proceso físico en condiciones ideales puede tener unos órdenes y unos coeficientes distintos. En cualquier caso, el Principio de Máxima Entropía garantiza que se está usando toda la información disponible y por tanto se está hallando la mejor aproximación del proceso que puede hallarse a partir de los datos disponibles. Esto es consistente con el Criterio de Corrección Científica que hemos introducido en la primera parte de la memoria.

4.3.2 Determinación de los coeficientes del modelo

El ajuste del modelo ARMA requiere determinar los coeficientes a_k, b_k que aparecen en la Eq. 4.3. Para determinar los parámetros de un modelo AR se puede utilizar un ajuste de mínimos cuadrados, el método de momentos o métodos Monte Carlo de Cadenas de Markov. Los coeficientes del modelo ARMA pueden obtenerse de igual manera pero es necesario, además, un procedimiento iterativo para obtener los coeficientes de la parte MA. La subrutina `armaint.m` usa un algoritmo iterativo conocido como [Steiglitz and McBride \(1965\)](#). Este algoritmo trata de minimizar el error cuadrático

$$\min_{a,b} \sum_{n=0}^{\infty} |x(n) - h(n)|^2 \quad (4.4)$$

donde $h(n)$ es la respuesta a impulso del filtro definido por $b(z)/a(z)$. El algoritmo comprueba que los coeficientes converjan después de un número de iteraciones predefinido. Dado que en lugar de minimizar el error de predicción o el error dado por una ecuación arbitraria como hacen otros métodos de predicción lineal, lo que minimiza este algoritmo es el error entre la serie, dada como respuesta a impulso de un filtro, y la propia respuesta a impulso que da el modelo, el método de Steiglitz-McBride da resultados considerablemente mejores.

Entre el gran número de algoritmos presentes en la literatura para la determinación de los coeficientes en modelos lineales existen algoritmos que resultan ser muy eficientes computacionalmente haciendo uso de una TFD ([Scargle, 1981](#)), sin embargo, se ha escogido el algoritmo de Steiglitz-McBride, además de por su óptimo rendimiento, para evitar el uso de transformadas de Fourier ya que requieren unas condiciones que, como se verá en la tercera parte de la memoria, no pueden darse por sentadas de partida. De este modo, el conjunto de funciones que pueden ser modeladas es mayor. La principal crítica que tiene este algoritmo es que sólo asume ruido blanco y de haber ruido coloreado trata este como si fuera ruido blanco. Esto podría afectar a la determinación de los coeficientes del modelo si lo que se trata es de excluir toda componente espectral que decae con una ley de potencias en frecuencias. Nuestro punto de partida es que todo aquello que no es ruido blanco contiene información y por tanto debe ser analizado, de modo que el algoritmo de Steiglitz-McBride sirve nuestros propósitos de forma correcta. Por otra parte, en las pruebas que presentamos aquí comparamos los resultados obtenidos con este métodos con los que se obtienen haciendo uso de otro método que sí que puede ajustar el ruido coloreado, este es el método de Prony ([Parks and Burrus, 1987](#)).

4. TÉCNICAS DE RELLENO DE HUECOS

4.3.3 Procedimiento de relleno

Para interpolar sobre el hueco el programa utiliza la subrutina `armafill.m`, que lleva a cabo el procedimiento haciendo un promedio pesado entre la predicción hallada a partir del modelo del segmento de datos anterior al hueco y la predicción hallada a partir del modelo del segmento de datos posterior al hueco. De este modo

$$\begin{aligned}x_n &= x_n^f w_n^f + x_n^b w_n^b, \\w_n^f &= 1 - w_n^b = w \cdot k, \quad k \in (1, l_g) \\w &= \frac{1}{1 + l_g}\end{aligned}\tag{4.5}$$

siendo x_n^f, x_n^b las predicciones por la izquierda y por la derecha, w_n^f, w_n^b los pesos que se aplican a dichas predicciones respectivamente, y l_g el tamaño del hueco.

El método fija un límite en el tamaño del hueco que puede rellenar, si este es mayor que el límite entonces itera el proceso retroalimentando las propias predicciones como datos para hacer la siguiente predicción de modo que se cubra todo el hueco a pesar del tamaño. Este límite debe ajustarse experimentalmente. Esto es algo fácil de hacer ya que la predicción empieza a perder la coherencia a partir de cierto valor y la amplitud cae considerablemente. Esto, más allá de ser un handicap del método puede interpretarse como una situación ventajosa ya que, a diferencia de otras técnicas que predicen valores independientemente de la fase, la pérdida de coherencia en este método asegura que sólo se predice la señal hasta donde se puede predecir su fase, de modo que puede cortarse la interpolación más allá de ese punto e iterar el algoritmo para seguir interpolando.

4.3.4 Flujo de datos

Describimos aquí el flujo de trabajo con los archivos de datos de CoRoT. Nótese que hay unas restricciones en cuanto al tamaño del hueco en relación al tamaño de los segmentos de datos que se usan para llevar a cabo la predicción. Esto debe ser así en todos los métodos de predicción ya que, en definitiva, no se pueden hacer predicciones de forma segura si no se imponen ciertas restricciones.

1. Las subrutinas `lincorr.m` y `sing.m` toman el vector de status de los datos de CoRoT de nivel 2 y realizan correcciones de forma que los segmentos de datos que se usarán posteriormente sean suficientemente significativos. De este modo se evitan errores en la interpolación por falta de datos. Por la misma razón se evita la interpolación en los extremos de la serie de datos.

2. `indgap.m` determina los índices de los huecos contenidos en la serie basándose en el status corregido de los datos. Se almacenan las posiciones de los huecos $n \in (N_q + 1, M_{q+1})$.
3. Se determina el orden (p, q) óptimo para el modelo mediante la subrutina `armaord.m`.
4. `armafill.m` selecciona los segmentos de datos locales que rodean el hueco q , es decir, $(M_q + 1, N_q)$ y $(M_{q+1} + 1, N_{q+1})$, calcula las predicciones por la izquierda y por la derecha en el hueco, y halla una predicción promedio utilizando una serie de pesos w_i que decaen con el inverso de la distancia.
5. Se repite el paso anterior para cada hueco introduciendo las predicciones calculadas y se vuelve al paso 2 al finalizar para hallar los nuevos índices de los huecos no interpolados en la iteración anterior.
6. Si vuelven a quedar los mismos huecos después de iterar desde el paso 2, se vuelve al paso 1 y se reconstruyen los índices de nuevo para poder rellenar nuevos huecos.
7. El algoritmo finaliza cuando no quedan huecos por rellenar o cuando las restricciones que se imponen para realizar la interpolación no permiten rellenar más huecos.

4. TÉCNICAS DE RELLENO DE HUECOS

Resultados: series sintéticas y curvas de luz de CoRoT

“...it follows that every movement in nature must be rhythmical...It is borne out in everything we perceive—in the movement of a planet, in the surging and ebbing of the tide, in the reverberations of the air, the swinging of a pendulum, the oscillations of an electric current, and in the infinitely varied phenomena of organic life. Does not the whole of human life attest to it? Birth, growth, old age, and death of an individual, family, race, or nation, what is it all but a rhythm? All life manifestation, then, even in its most intricate form, as exemplified in man, however involved and inscrutable, is only a movement, to which the same general laws of movement which govern throughout the physical universe must be applicable.”

Nikola Tesla

5.1 Tests del método mediante simulaciones numéricas

En esta sección realizamos los tests del método de interpolación sobre series sintéticas generadas a partir de distintos tipos de modelos según el tipo de función base que se utiliza en cada caso: función periódica sinusoidal con distintos rangos de frecuencias - por definición analíticas, y función que es por construcción no-analítica. El primer tipo es el modelo estándar que se supone usualmente cuando se reconstruye la señal debida a pulsaciones estelares. Para la evaluación de los resultados se realiza un análisis de frecuencias con un nuevo código desarrollado (ver aplicaciones de este código en [Uytterhoeven et al. 2011](#); [Pascual-Granado et al. 2013](#)) que pretende ser coherente con el criterio objetivo en el que nos basamos y es, además, computacionalmente más eficiente que otros códigos usados habitualmente en la literatura de astrosismología ([Reegen 2007](#); [Lenz and Breger 2005](#)). Este código se describe en los Apéndices (ver Apéndice [A](#)).

5.1.1 Función periódica sinusoidal

El modelo usado en este caso es una suma de 2 términos: el primero, $\epsilon(t)$, que corresponde a un ruido, y el segundo corresponde a la señal sinusoidal. La primera componente se suele considerar generalmente ruido blanco con distribución gaussiana pero aquí probaremos que el método de interpolación ARMA no es sensible al tipo de ruido ya que es capaz de modelar las variaciones del ruido coloreado también. Por otro lado, la componente de señal puede representarse como:

$$X(t) = \sum_{n=1}^{n_f} A_n \cos(2\pi f_n t + \phi_n) \quad (5.1)$$

donde n_f es el número de componentes sinusoidales de la señal, y A_n, f_n, ϕ_n representan la amplitud, frecuencia, y fase de la componente n -ésima de la señal.

En los dos primeros tests, cuyos resultados presentamos más abajo, analizamos la respuesta del método según el rango de frecuencias considerado. Para ello elegimos $n_f = 3$ para simplificar la evaluación de los resultados. Posteriormente, en el tercer test, analizamos un modelo de una curva de luz de una estrella con un elevado número de frecuencias de pulsación.

Para el análisis utilizamos una discretización del modelo presentado arriba con un tiempo de muestreo $\Delta t = 32$ s, que se corresponde con las 32 integraciones de 1 segundo que realiza CoRoT para el seismocampo; el tiempo de observación es de $T = 30$ días, que es aproximadamente la duración de un *Short-Run*, lo que hace una serie de $N = 81000$ puntos.

Por otra parte, para generar los huecos en la serie utilizamos la ventana del LRC02 de CoRoT tomando los 30 días centrales de la serie (ver Fig. 5.1). Puede apreciarse la regularidad de los valores distintos de cero correspondientes en su mayor parte al paso por la Anomalía Suratlántica, que está identificado con los valores 12 y 4 del status según que las medidas hayan sido interpoladas o no con el método de interpolación lineal que incorpora el pipeline de CoRoT a los datos de nivel 2. En total se han eliminado 7614 datos de la serie lo que hace aproximadamente un 10% de los datos. El test que realizamos comprueba la eficiencia del método de interpolación al reconstruir las características espectrales de la señal $X(t)$ a partir de la serie que resulta del producto $X(t_j) \cdot W(t_j)$, donde $W(t_j)$ es la función ventana dada por la ventana de CoRoT.

5.1.1.1 Bajas frecuencias

En este apartado llevamos a cabo el test del método de interpolación cuando se hace uso de éste para rellenar huecos de una serie que se corresponde a una señal con variaciones de bajas frecuencias. Hablamos aquí de bajas o altas frecuencias

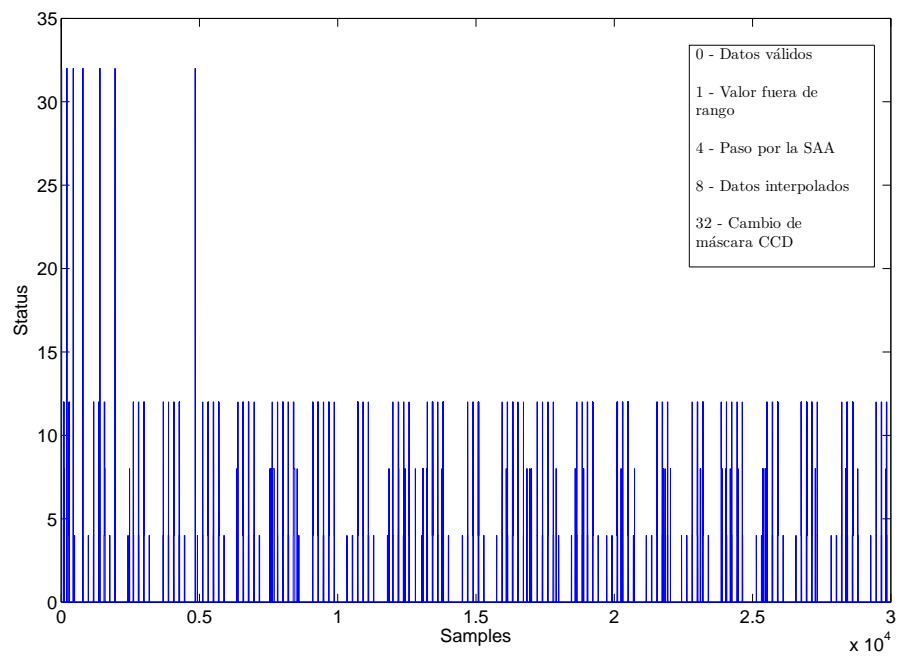


Figura 5.1: Ventana utilizada para generar los huecos en los tests a partir del status de los datos observados por CoRoT en el LRc02. En ordenadas se representan los valores del status de los datos, que toman valor cero cuando no se ha producido ningún evento durante la observación. Nótese la periodicidad introducida por los pasos a través de la Anomalía Suratlántica.

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

A_n	f_n (cd ⁻¹)	ϕ_n (rads)
1.00	1.257	0.2719
0.66	2.392	4.1628
0.40	1.363	1.3532

Tabla 5.1: Parámetros del modelo usado para el test de bajas frecuencias del método de interpolación ARMA donde A_n son amplitudes, f_n son las frecuencias, y ϕ_n las fases

en relación a la frecuencia de muestreo ($f_s = 2700$ d⁻¹). Así pues nos referimos a frecuencias altas cuando éstas son $f \geq f_s/20 = 135$ d⁻¹ y nos referimos a frecuencias bajas cuando éstas son $f \leq f_s/200 = 13.5$ d⁻¹. En este caso, haciendo uso de los parámetros indicados anteriormente, el modelo discreto tiene la siguiente expresión:

$$X_j = A_1 \cos(2\pi f_1 j \Delta t + \phi_1) + A_2 \cos(2\pi f_2 j \Delta t + \phi_2) + \quad (5.2)$$

$$+ A_3 \cos(2\pi f_3 j \Delta t + \phi_3) + \varepsilon_{j\Delta t}, \quad \text{con } j = 1, \dots, N$$

Los parámetros utilizados en este caso se recogen en la tabla 5.1. El término $\varepsilon_{j\Delta t}$, que corresponde a ruido blanco, se genera a partir de una distribución gaussiana con media cero y desviación estándar igual a 1/10 de la amplitud A_n máxima utilizada para la señal. Utilizamos $j\Delta t$ como subíndice para destacar que ε no es una función del tiempo ya que es una componente estocástica. Con los valores asignados a los parámetros anteriores se cumple que el coeficiente de señal-ruido de la señal es $\text{SNR} > 4$, que es el valor mínimo sugerido por R. Garrido para la detección de señal en [Breger et al. \(1993\)](#). En la fig. 5.2 puede verse la serie completa del modelo analítico de bajas frecuencias que utilizamos para este primer test del método de interpolación ARMA.

Ahora, tras realizar el producto de X_j con la función ventana, para realizar el relleno ARMA el primer paso es determinar el orden óptimo para el modelo ARMA que se utilizará para generar las predicciones de la serie en los huecos. Para ello, como fue mencionado anteriormente, el algoritmo determina el modelo ajustado cuyo coeficiente de Akaike (AIC) es menor de entre un conjunto de modelos con órdenes en un rango que va desde el (2,0) hasta el (20,20). Según el contenido de la señal de la serie el coeficiente de Akaike puede variar con el número de datos del segmento modelado hasta un valor de N a partir del cual converge. Por tanto, para asegurarnos de seleccionar el orden óptimo tomamos el mayor segmento sin huecos de la serie ($N = 640$ puntos) para estimar el orden óptimo. En la fig. 5.3 se han representado los valores del coeficiente de Akaike (AIC) en función de los órdenes p y q correspondientes a la parte AR y MA respectivamente del modelo ARMA. Para visualizar los resultados los valores AIC se han normalizado y se ha tomado el logaritmo, de modo que el

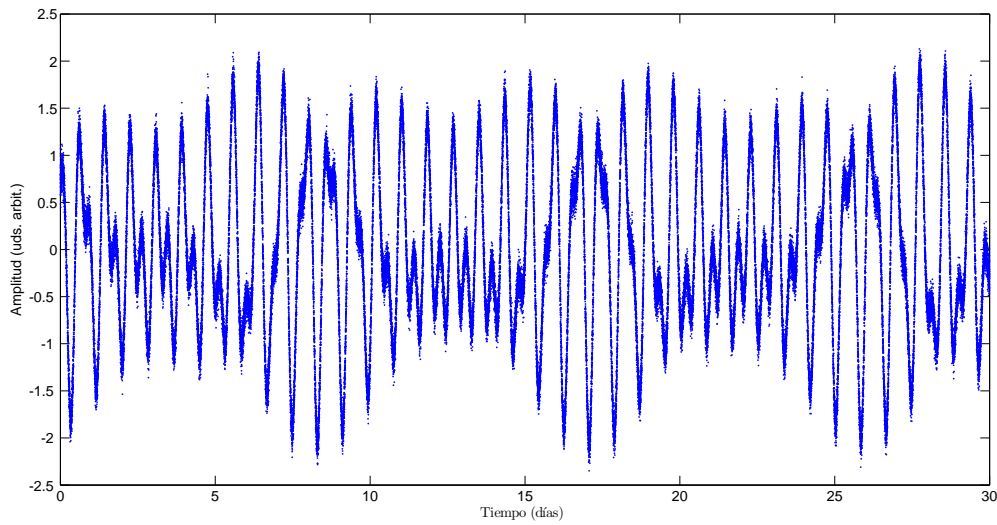


Figura 5.2: Realización de un modelo analítico compuesto por una señal de bajas frecuencias y ruido blanco gaussiano.

mínimo absoluto de la matriz de valores de AIC obtenida sea cero. En el caso mostrado en la figura el mínimo absoluto se encuentra para el modelo ARMA(8,4) y el siguiente mínimo se halla para el modelo ARMA(4,8).

Es fácil ver que un modelo AR(2) puede representar un proceso armónico ya que este puede ser representado a su vez mediante una ecuación en diferencias de segundo orden que tiene una expresión similar a la del modelo AR(2) (Ulrych and Ooe, 1979). Por otro lado, como se mencionó en § 4.3.1, un modelo ARMA(p,p) es la representación más parsimoniosa de un proceso armónico con ruido aditivo. De modo que, si analizamos una señal con $p/2$ componentes armónicas, el modelo más parsimonioso que representa una realización de esta señal en condiciones ideales es el ARMA(p,p). En el caso que nos toca, una señal que contiene 3 componentes armónicas tiene su representación ideal en forma de proceso autoregresivo en un modelo ARMA(6,6). Puesto que las condiciones no son ideales, es decir, el tamaño del segmento modelado es finito y el ruido no es despreciable, el mínimo absoluto de AIC no se encuentra en el orden (6,6) sino en el (8,4), pero el criterio de Akaike garantiza que este es el modelo óptimo bajo las condiciones del análisis, es decir, con la información contenida en los datos del segmento modelado. Puede observarse, además, que el modelo ARMA(6,6) se halla entre los 2 primeros mínimos locales (8,4) y (4,8) probando así que este criterio proporciona una buena aproximación a la representación en condiciones ideales.

Una vez determinado el orden óptimo es posible ajustar el modelo a los segmentos de datos que preceden y prosiguen los huecos para realizar las predicciones hacia delante y hacia atrás en cada hueco. En la Fig. 5.4 comparamos los datos interpolados linealmente (panel medio) y mediante ARMA (panel inferior) de un segmento de 0.75

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

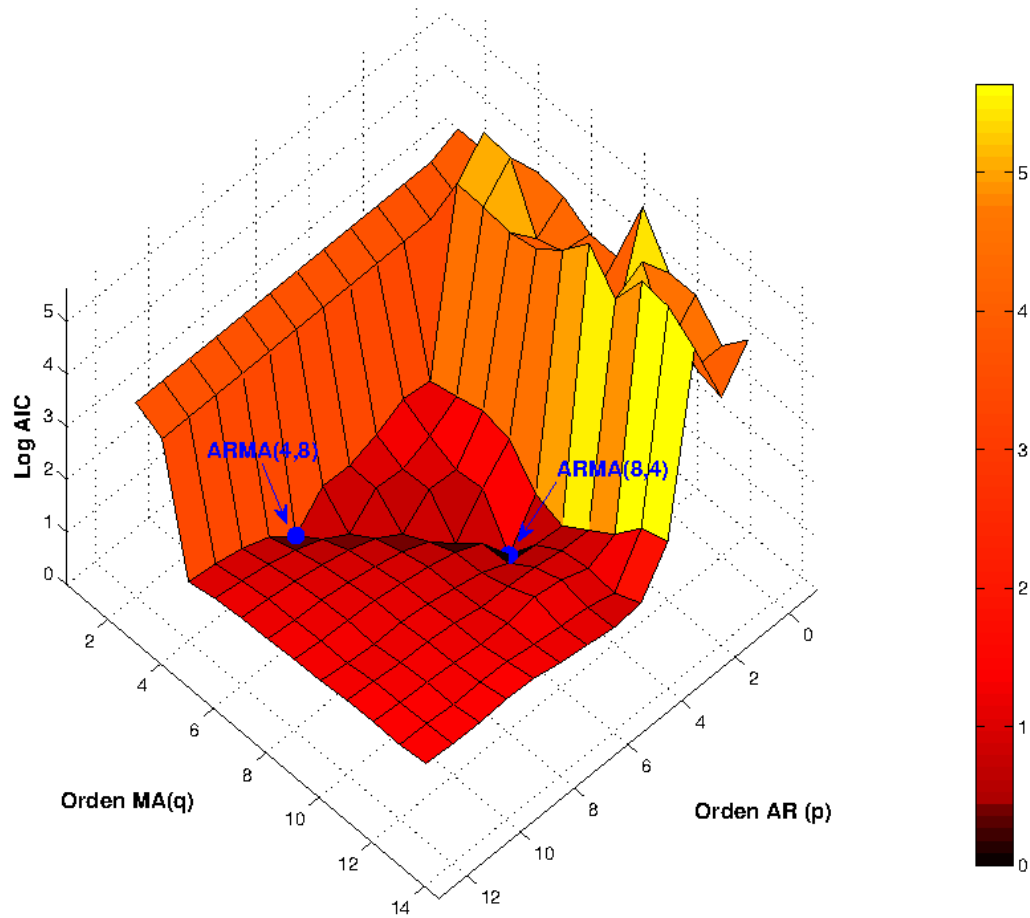


Figura 5.3: Estimación del orden ARMA(p,q) para el modelo que se utiliza en la reconstrucción de los huecos de la serie sintética compuesta por una señal de bajas frecuencias y ruido blanco gaussiano. Nótese como se alcanzan mínimos en (8,4) y (4,8).

días de la serie original que aparece en el panel superior sin huecos. A diferencia de la interpolación lineal el método ARMA recupera la pendiente de la serie en los huecos de modo que no introduce ningún error en las derivadas y la señal correspondiente a las variaciones de la serie es preservada. Por otro lado, los segmentos interpolados aparecen diferenciados por el hecho de que sólo se reconstruye la señal y no el ruido, es decir, el ruido en la serie interpolada es menor que en la serie original. Esto, lejos de ser un inconveniente es una ventaja para el análisis de señal.

Las observaciones comentadas en el párrafo anterior, sin duda, pueden apreciarse mejor al analizar la señal en el dominio de frecuencias. En las Figs. 5.5 y 5.6 hemos comparado en escala logarítmica el espectro que se obtiene a partir de la serie con huecos con el obtenido al interpolar los huecos linealmente y con el método ARMA respectivamente. Con respecto a la determinación de las frecuencias principales de la señal, usadas para el test de señal con bajas frecuencias, la precisión es la misma tanto para la serie interpolada linealmente como para la serie interpolada con ARMA. Sin embargo, las amplitudes de estas 3 frecuencias son distintas, siendo en torno a un 0.2% menores en el caso de la serie interpolada linealmente. En el resto del rango de frecuencias se observa que ambos métodos reducen considerablemente el efecto de la ventana espectral mostrado en la figura en azul. No obstante, para apreciar las diferencias entre ambos métodos de interpolación en todo el rango de frecuencias estudiado en la Fig. 5.7 hemos representado el periodograma de la serie sin huecos junto a los dos periodogramas de las series interpoladas y en el panel inferior el cociente entre ambas. Como puede observarse a través del inserto y del panel inferior de la figura, las diferencias entre el periodograma de la serie interpolada linealmente y la serie interpolada mediante ARMA empiezan a hacerse más notables a partir de 6 d^{-1} con amplitudes mayores en el caso de la primera. Esto implica que la interpolación lineal, si bien permite determinar de forma correcta las frecuencias principales y en menor medida sus amplitudes, no corrige los alias que aparecen a lo largo del espectro debido a la ventana espectral y el fenómeno de *leakage*.

5.1.1.2 Altas frecuencias

Los parámetros utilizados en este caso se recogen en la tabla 5.2 y el $\varepsilon_{j\Delta t}$ se ha calculado por el procedimiento anterior. En la fig. 5.8 puede verse un segmento de 1 día de la serie que hemos generado a partir del modelo analítico de altas frecuencias para el segundo test del método de interpolación ARMA.

En cuanto a la estimación del orden, en este caso, como se muestra en la Fig. 5.9, se requiere un mayor número de coeficientes para representar la serie (28 en total sumando la parte AR y la parte MA). Esto es debido a que al ser las frecuencias mayores que en el caso anterior las variaciones están peor muestreadas y se requieren

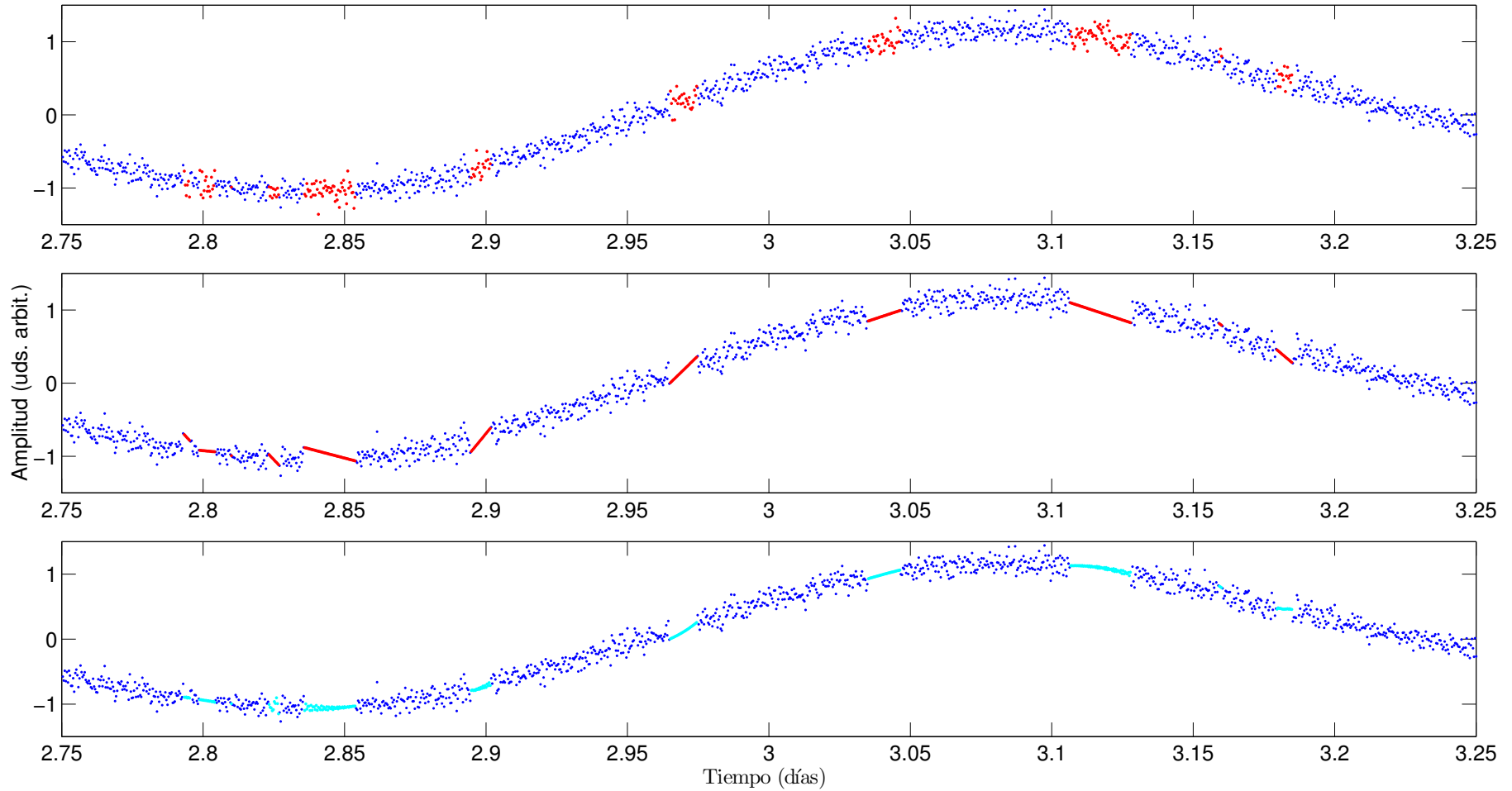


Figura 5.4: Comparación entre el método ARMA e interpolación lineal para el relleno de huecos. En el panel superior se muestra un segmento de 0.75 días de la serie generada con el modelo analítico de bajas frecuencias. Los segmentos marcados en rojo son los que han sido interpolados en el panel central (interpolación lineal) y en el inferior (interpolación ARMA).

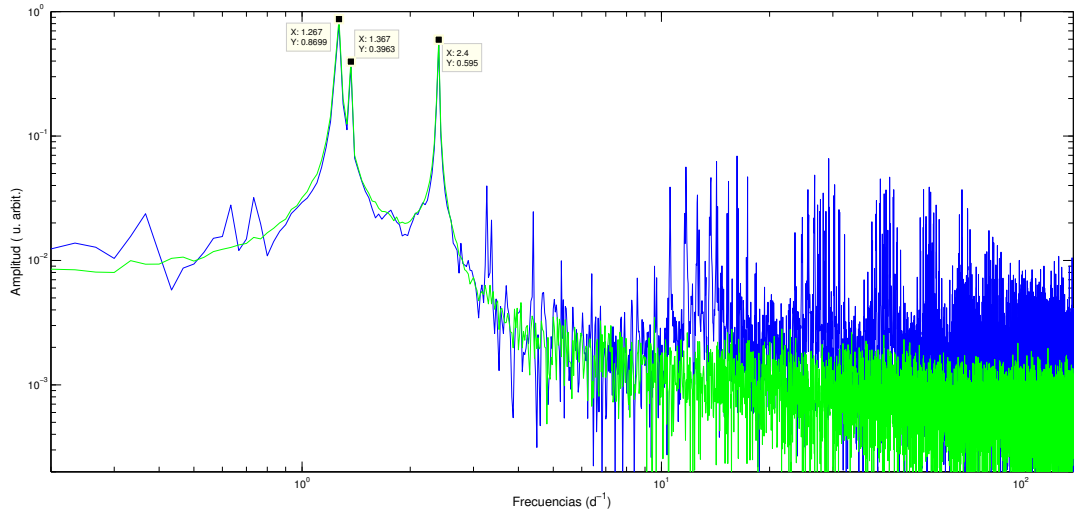


Figura 5.5: Periodograma en escala logarítmica de la serie interpolada linealmente (en verde) comparada con el periodograma de la serie con huecos (azul).

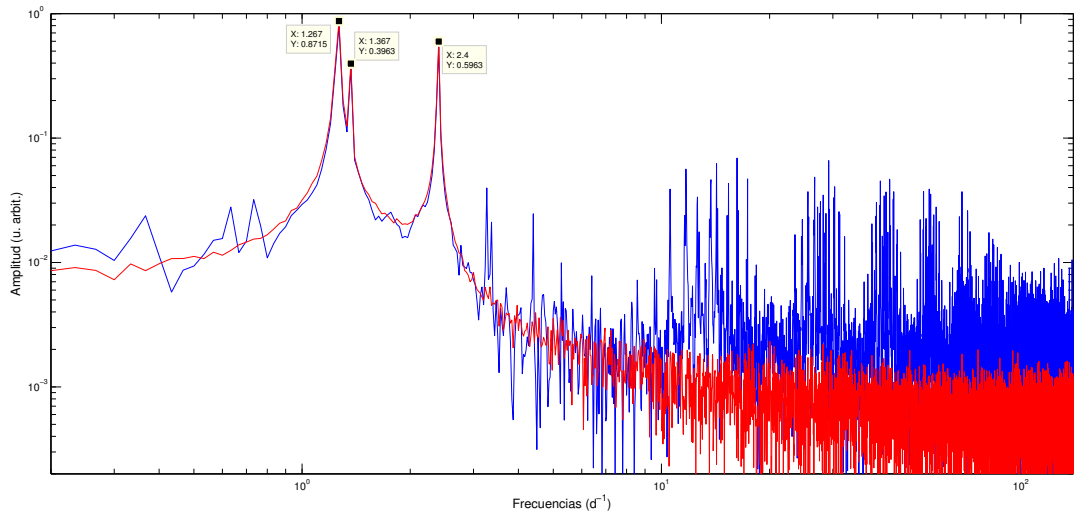


Figura 5.6: Periodograma en escala logarítmica de la serie interpolada con ARMA (en rojo) comparada con el periodograma de la serie con huecos (azul). Nótese que en este caso las amplitudes de los frecuencias principales de la señal son mayores que en la figura anterior.

A_n	f_n (cd^{-1})	ϕ_n (rads)
1.00	181.723	0.2719
0.66	172.392	4.1628
0.40	214.631	1.3532

Tabla 5.2: Parámetros del modelo usado para el test de altas frecuencias del método de interpolación ARMA donde A_n son amplitudes, f_n son las frecuencias, y ϕ_n las fases

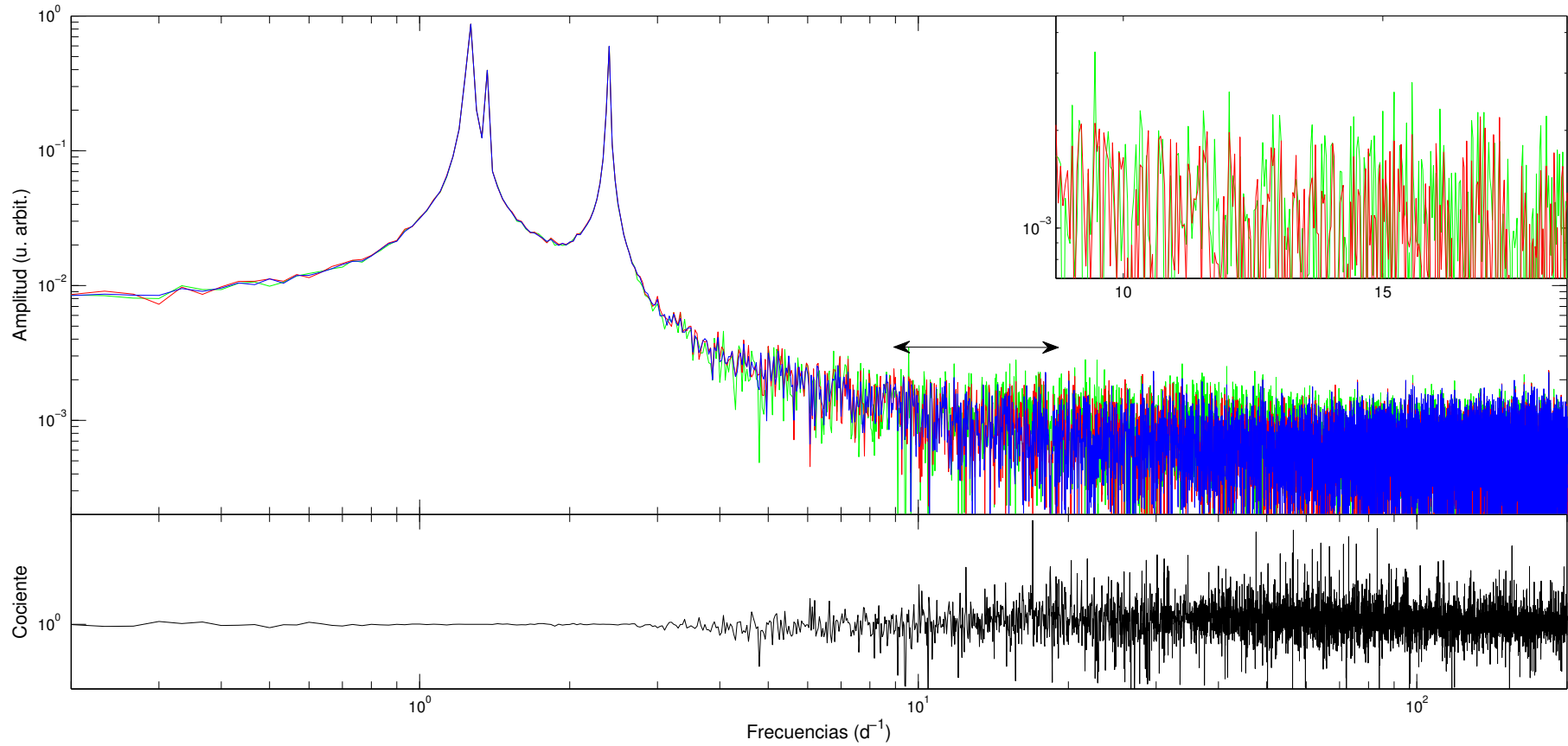


Figura 5.7: Panel superior: Periodograma en escala logarítmica de la serie sin huecos (en azul) comparada con el periodograma de la serie interpolada con ARMA (en rojo) y el correspondiente a la serie interpolada linealmente (en verde). En el inserto se muestran estas dos últimas en el rango marcado por una flecha de dos puntas que se corresponde a las frecuencias de 9 a 20 d^{-1} . Panel inferior: Cociente entre las amplitudes obtenidas para la serie interpolada linealmente y las que se obtienen al interpolar los huecos con ARMA. Nótese como las diferencias entre ambos periodogramas se hacen significativas a partir de los 6 d^{-1} .

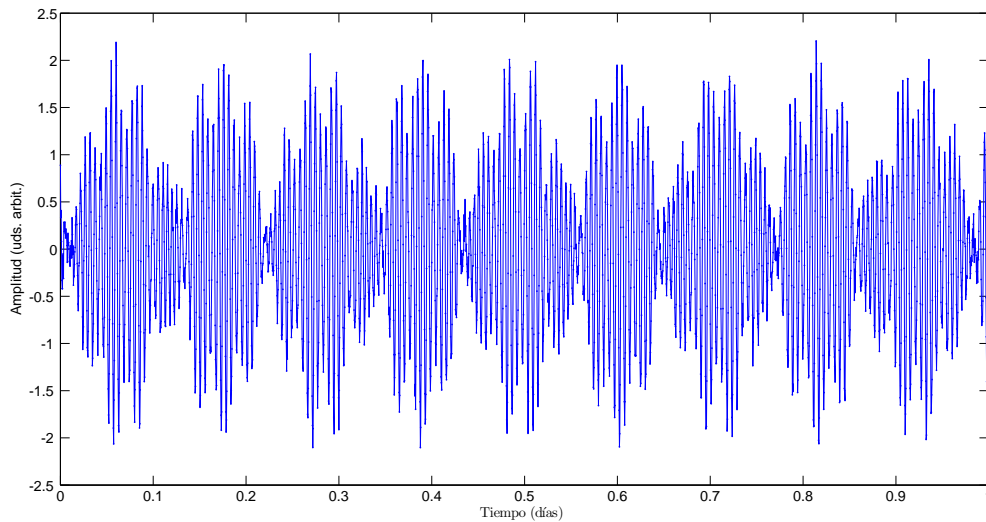


Figura 5.8: Realización de un modelo analítico compuesto por una señal de altas frecuencias y ruido blanco gaussiano.

mejores condiciones de ruido y/o un segmento de una longitud mayor para poder representar la señal con un $\text{ARMA}(6,6)$. No obstante, como se indicó anteriormente, el criterio de Akaike nos asegura que el orden $(21,7)$ es el óptimo para representar la señal en las condiciones simuladas.

A continuación ajustamos el modelo $\text{ARMA}(21,7)$ a los segmentos de datos y hallamos las predicciones que permiten rellenar los huecos de la serie estudiada. En la Fig. 5.10 comparamos el relleno de huecos mediante el método de interpolación lineal (panel central) y mediante ARMA (panel inferior) del segmento de 0.5 días de la serie original que aparece en el panel superior sin huecos. En este caso, puede observarse que la interpolación lineal destruye por completo la señal en los huecos mientras que el método ARMA reconstruye los segmentos marcados en rojo en el panel superior de forma idéntica preservando por completo la señal pese a que el orden $(21,7)$ del modelo no es el ideal.

Al igual que en el test anterior, en la Fig. 5.11 hemos comparado con el periodograma de la serie con huecos en escala logarítmica, el periodograma de la serie interpolada linealmente, y el de la serie interpolada mediante el método ARMA, en la Fig. 5.12. En este caso vemos que el espectro obtenido tras hacer una interpolación lineal contiene en el rango de bajas frecuencias alias de mayor amplitud incluso que el espectro de la serie con huecos. Es decir, en este caso, el efecto de poner ceros en los huecos es menos contraproducente para el análisis espectral que hacer una interpolación lineal. Por otro lado, en la Fig. 5.12 se puede apreciar que los lóbulos que aparecen en torno a los picos de las frecuencias principales se reducen considerablemente al interpolar mediante el método ARMA. En la Fig. 5.13 puede apreciarse

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

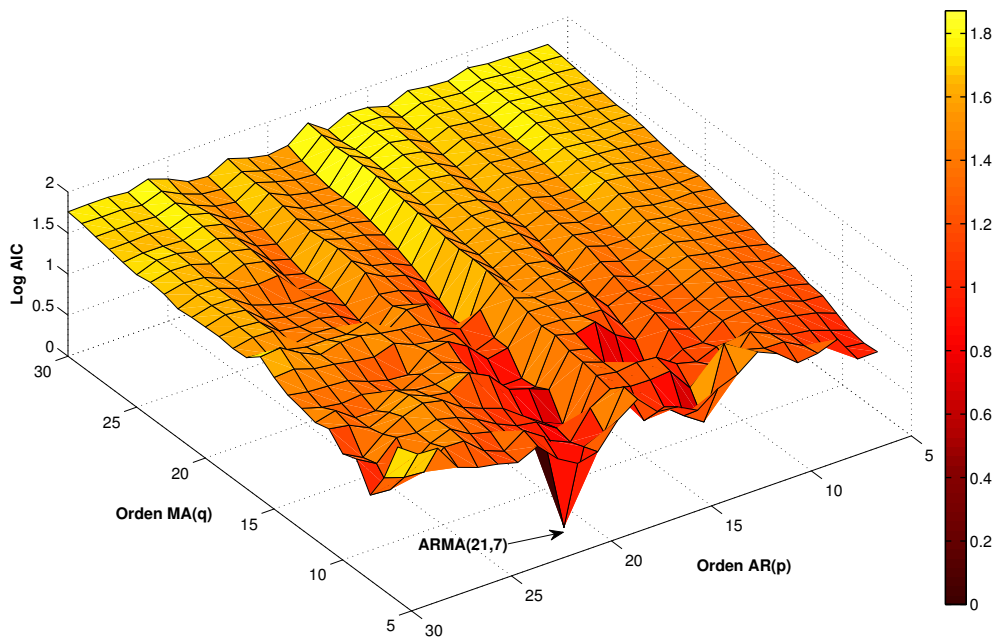


Figura 5.9: Estimación del orden ARMA(p,q) para el modelo que se utiliza en la reconstrucción de los huecos de la serie sintética compuesta por una señal de altas frecuencias y ruido blanco gaussiano.

también, que la amplitud de los alias es mucho menor en todo el rango de frecuencias cuando se interpola con ARMA que al interpolar la serie linealmente.

Para calcular numéricamente las diferencias en la determinación de los parámetros de cada frecuencia en cada caso hacemos un ajuste no-lineal de mínimos cuadrados con las frecuencias fijas y la amplitud y fase variables. Las frecuencias que usamos para estos ajustes son las que hemos identificado previamente a partir de los picos de mayor amplitud de los espectros. Estos ajustes nos dan los resultados mostrados en las tablas 5.3 y 5.4. Es notable que los valores de las amplitudes que aparecen en la tercera y cuarta columna difieren mínimamente, es decir, en lo que a determinación de amplitudes se refiere, se obtiene resultados similares al interpolar linealmente a los que se obtiene al no interpolar. Por otro lado, puede apreciarse que, mientras que las diferencias entre las amplitudes de la serie interpolada linealmente A_n^{lin} con las amplitudes de la serie sin huecos A_n^0 están en torno al 10%, estas no llegan al 1% en el caso de las amplitudes de la serie interpolada con ARMA. Puede notarse, también, que las fases determinadas a partir de la serie interpolada con ARMA se acercan más a las originales que las fases determinadas a partir de la serie interpolada linealmente. No obstante, las diferencias en la determinación de las fases no son tan significativas en este caso como las diferencias en la determinación de las amplitudes, que es el efecto más determinante en la detección de frecuencias.

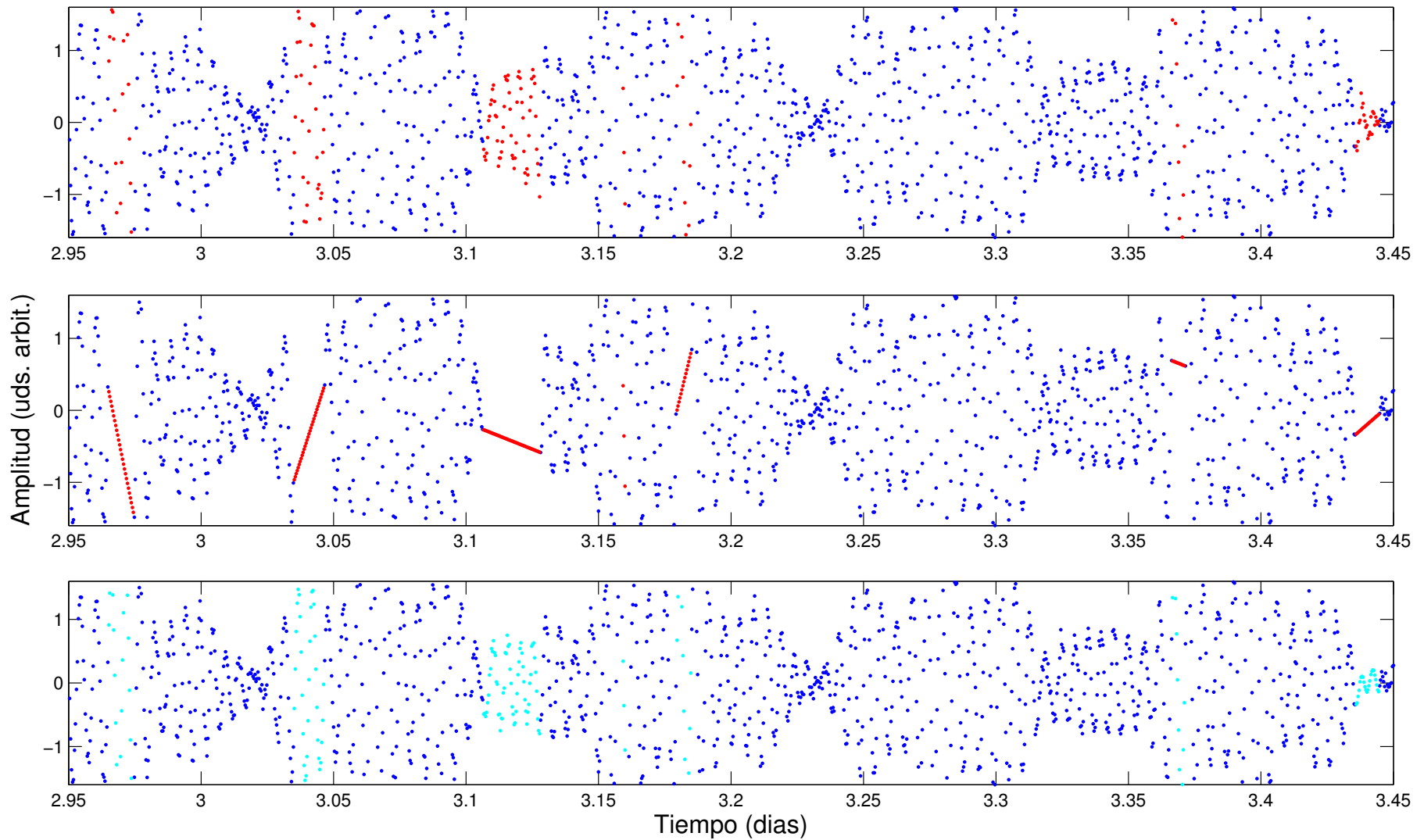


Figura 5.10: Comparación entre el método ARMA para el relleno de huecos e interpolación lineal. En el panel superior se muestra un segmento de 0.5 días de la serie generada con el modelo analítico de altas frecuencias. Los segmentos marcados en rojo son los que han sido interpolados en el panel central (interpolación lineal) y en el inferior (interpolación ARMA).

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

f_n (cd^{-1})	A_n^0	A_n^{gap}	A_n^{lin}	A_n^{ar}
181.7233	0.9995 ± 0.0005	0.9058 ± 0.0013	0.9058 ± 0.0018	0.9906 ± 0.0006
172.3933	0.6578 ± 0.0005	0.5958 ± 0.0013	0.5963 ± 0.0018	0.6636 ± 0.0006
214.6300	0.3992 ± 0.0005	0.3616 ± 0.0013	0.3614 ± 0.0018	0.3955 ± 0.0006

Tabla 5.3: Amplitudes estimadas mediante un ajuste de mínimos cuadrados no-lineal a partir de la serie generada con el modelo de frecuencias altas sin huecos A_n^0 , con huecos A_n^{gap} , con interpolación lineal A_n^{lin} , y con interpolación ARMA A_n^{ar} .

f_n (cd^{-1})	φ_n^0	φ_n^{gap}	φ_n^{lin}	φ_n^{ar}
181.7233	0.2442 ± 0.0005	0.2431 ± 0.0015	0.2429 ± 0.0020	0.2449 ± 0.0006
172.3933	4.0396 ± 0.0008	4.0409 ± 0.0023	4.0409 ± 0.0030	4.0404 ± 0.0009
214.6300	1.4488 ± 0.0013	1.4559 ± 0.0038	1.4558 ± 0.0050	1.4472 ± 0.0016

Tabla 5.4: Fases estimadas mediante un ajuste de mínimos cuadrados no-lineal a partir de la serie generada con el modelo de frecuencias altas sin huecos φ_n^0 , con huecos φ_n^{gap} , con interpolación lineal φ_n^{lin} , y con interpolación ARMA φ_n^{ar} .

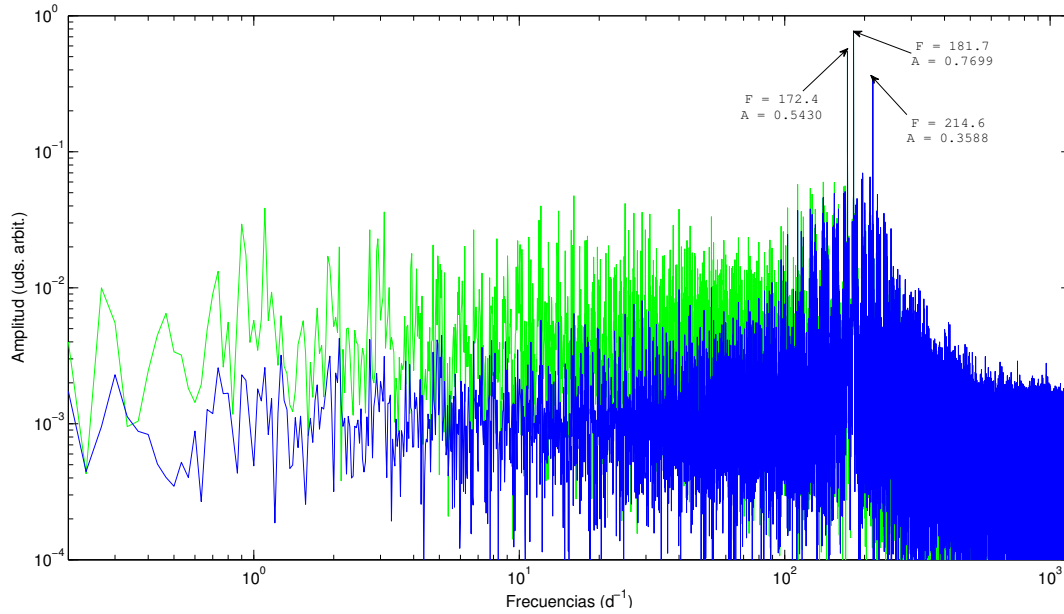


Figura 5.11: Periodograma en escala logarítmica de la serie interpolada linealmente (en verde) comparada con el periodograma de la serie con huecos (azul).

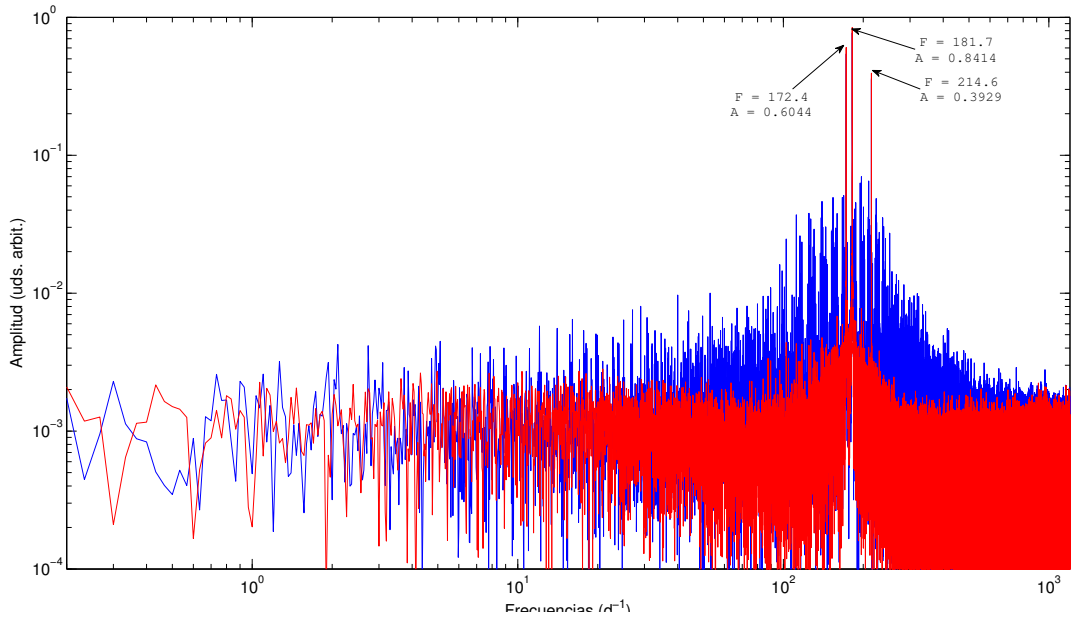


Figura 5.12: Periodograma en escala logarítmica de la serie interpolada con ARMA(en rojo) comparada con el periodograma de la serie con huecos (azul). Nótese que en este caso las amplitudes de los frecuencias principales son en torno a un 10% mayores que en la figura anterior.

5.1.1.3 Modelo realista

En los tests anteriores hemos estudiado los efectos del método de interpolación sobre la determinación de las frecuencias, amplitudes y fases según sea el rango de las frecuencias que pueden extraerse de la serie temporal. En esos casos, para simplificar la situación y poder apreciar los efectos con más detalle se usaron sólo 3 frecuencias. En el caso que ahora nos ocupa, el modelo usado para generar las series es similar al de la Eq. 5.1 pero con un número de frecuencias mucho mayor $n_f = 31$. Además, ahora la lista de frecuencias no es arbitraria sino que un conjunto de ellas se ordenan siguiendo el patrón dado por la fórmula de Tassoul (1980) para el régimen asintótico, que describimos a continuación.

Esta fórmula, que se deduce de la teoría de pulsación estelar, nos dice que las frecuencias de orden radial n alto y grado azimutal ℓ bajo (típicamente de 0 a 3) siguen un patrón dado por la ecuación:

$$\nu_{\ell,n} \simeq \left(n + \frac{\ell}{2} + \frac{1}{4} + \alpha \right) \Delta\nu \quad (5.3)$$

donde α es una constante de fase, y $\Delta\nu$ es un espaciamiento constante entre frecuencias que se conoce como “gran separación”. Este espaciamiento puede determinarse directamente a partir de

$$\Delta\nu = \nu_{n+1,\ell} - \nu_{n,\ell} \quad (5.4)$$

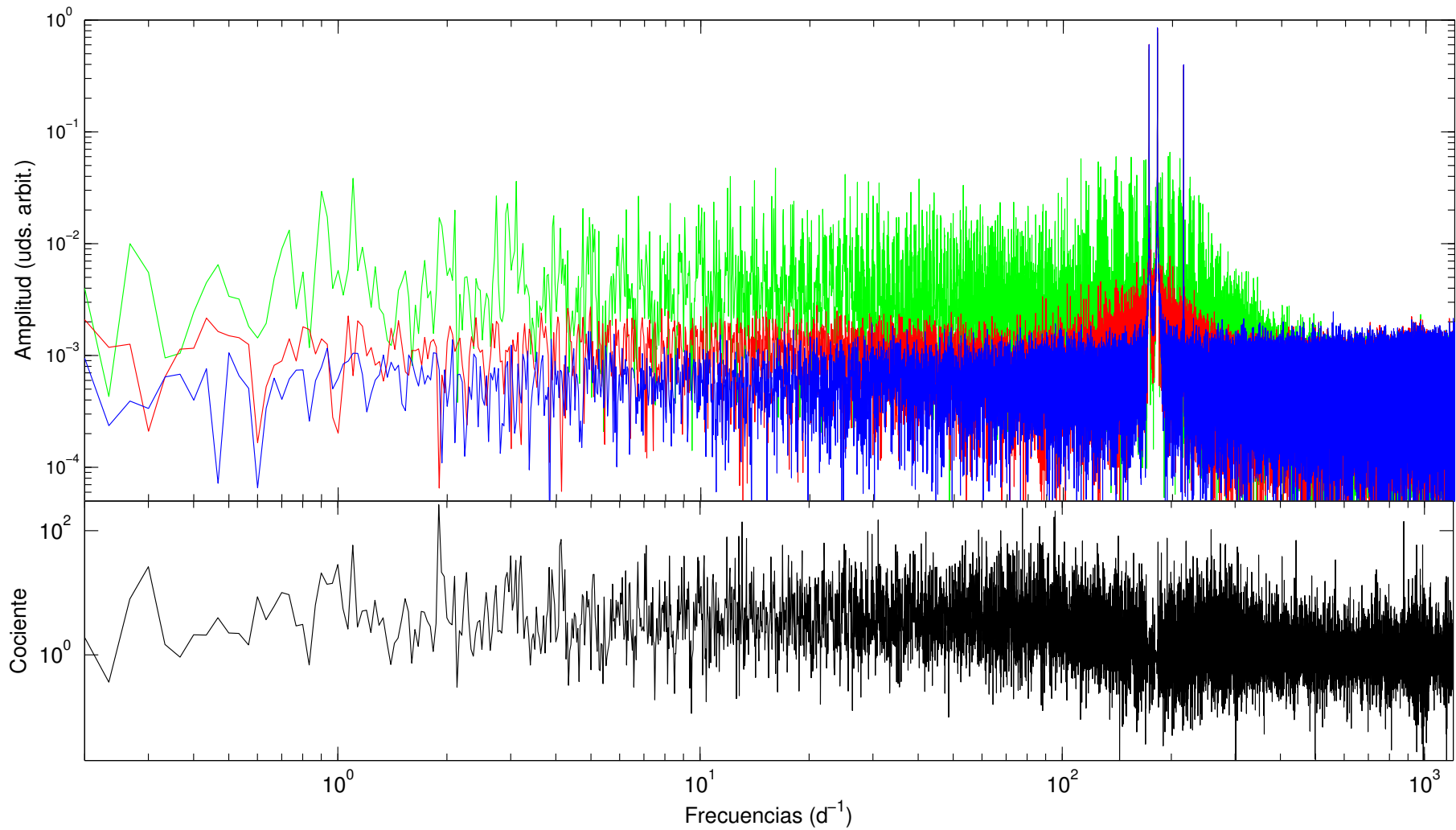


Figura 5.13: Panel superior: Periodograma en escala logarítmica de la serie sin huecos (en azul) comparada con el periodograma de la serie interpolada con ARMA(en rojo) y el correspondiente a la serie interpolada linealmente (en verde). Panel inferior: Cociente entre las amplitudes obtenidas para la serie interpolada linealmente y las que se obtienen al interpolar los huecos con ARMA. Nótese como las diferencias entre ambos periodogramas son significativas en todo el rango por debajo de los 400 d⁻¹.

y está relacionado con la densidad media en la estrella lo que hace que sea un parámetro de enorme interés en física estelar. Por otra parte, de la Eq. 5.3 se deduce una segunda separación constante conocida como “pequeña separación”:

$$\delta\nu = \nu_{n,\ell} - \nu_{n-1,\ell+2} \simeq (4\ell + 6)D_0 \quad (5.5)$$

es decir, los modos de igual $n + \ell$ están cuasi degenerados. En la Eq. 5.5 D_0 es una constante que viene determinada por las condiciones en el núcleo estelar. Con estas 2 separaciones constantes, la segunda típicamente mucho menor que la primera, se obtienen grupos separados por una distancia en frecuencias de $\Delta\nu/2$ correspondientes a los grados ℓ pares e impares.

En el modelo que usamos para generar las series temporales de este test la fórmula de Tassoul se implementa de la siguiente manera:

$$\begin{aligned} \nu_{\ell,n} &= \left(n + \frac{\ell}{2} + \frac{1}{4} + \alpha \right) \Delta\nu & \ell = 0, 1 \\ \nu_{\ell,n} &= \left(n + \frac{\ell}{2} + \frac{1}{4} + \alpha \right) \Delta\nu + \delta\nu & \ell = 2, 3 \end{aligned} \quad (5.6)$$

donde n puede tomar valores de 10-16, $\alpha = 0$, $\Delta\nu = 50 \mu\text{Hz}$ y $\delta\nu$ viene dado por la Eq. 5.5 con $D_0 = 0.16 \mu\text{Hz}$ de modo que para $\ell = 0$ resulta $\delta\nu = 1 \mu\text{Hz}$. Esto nos da un patrón de 28 frecuencias entre los $512.5 \mu\text{Hz}$ y los $888.5 \mu\text{Hz}$ cuyas amplitudes se distribuyen en una campana de Gauss centrada en $700 \mu\text{Hz}$.

Este patrón es típico de estrellas de tipo solar y aparece también en el espectro de gigantes rojas pero, en el modelo que estamos utilizando las frecuencias son estables y sus amplitudes no varían durante los 30 días de duración de la serie temporal, y el mecanismo de excitación de los modos en estrellas de tipo solar o gigantes rojas es de tipo estocástico, por lo que tienen un tiempo de vida Γ que puede ser del orden de la duración de las observaciones. De modo que, el modelo que estamos utilizando se ajustaría mejor a pulsaciones excitadas por el mecanismo de opacidad. Este es el mecanismo predominante en estrellas δ Sct pero, dado que la fórmula de Tassoul describe una tendencia asintótica cuando el orden radial $n \rightarrow \infty$, y en estrellas δ Sct las pulsaciones ocurren en torno al modo fundamental radial, no es de esperar que este patrón aparezca en su espectro. No obstante, desde 1997 en que se atribuyó un espaciado en frecuencias de $20 \mu\text{Hz}$ en el espectro de la estrella CD-247599 citepHAN97 a la gran separación, se ha estudiado la posibilidad de que exista este patrón en varias estrellas δ Sct. Así pues, el modelo que utilizamos para el test de esta sección se asemeja al que, según la teoría de pulsación estelar y las observaciones más recientes, se espera que pueda tener una estrella δ Sct.

Las otras 3 frecuencias que no siguen la fórmula de Tassoul se distribuyen como el fundamental radial y los 2 primeros sobretonos de acuerdo a las relaciones de [Petersen](#)

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

(1973):

$$\frac{\nu_0}{\nu_1} = 0.77 \quad \frac{\nu_0}{\nu_2} = 0.62 \quad (5.7)$$

donde $\nu_0 = 10 \mu\text{Hz}$. Con las dos expresiones anteriores obtenemos un modelo que permite generar una serie temporal similar a la de las curvas de luz de estrellas pulsantes δ Sct, y es por ello que hemos llamado “realista” a este modelo, sin embargo, se trata simplemente de una aproximación útil para nuestro objetivo, que no es sino estimar los efectos de la interpolación sobre la determinación de las frecuencias. Si quisiéramos generar una aproximación mejor a las curvas de luz de las δ Sct deberíamos utilizar un desarrollo más complejo para determinar la lista de frecuencias partiendo de trabajos como el de [Roxburgh and Vorontsov \(1994\)](#), lo que daría lugar a una relación entre frecuencias que no fuera exactamente constante.

La serie temporal generada a partir del modelo expuesto arriba nos permitirá comprobar otros aspectos de la interpolación en los huecos no tenidos en cuenta con anterioridad como son:

- El efecto de aumentar el número de frecuencias de la señal sobre la interpolación. Es decir, si el método de interpolación podría preservar la información en una situación crítica.
- Cómo afecta a la determinación de las fases la cuasi degeneración de los modos que están separados únicamente por $\delta\nu$.
- El efecto del método de interpolación sobre la determinación de la gran separación usando la técnica de la transformada de Fourier segunda ([García Hernández, 2011](#)).

Estudiamos estos aspectos en un modelo con el tipo de señal expuesta y una componente estocástica con una distribución espectral de ruido blanco.

En la Fig. 5.14 se representa, como anteriormente, la matriz de coeficientes de Akaike para hallar el orden óptimo del modelo ARMA que representa la serie temporal generada. El modelo que minimiza el error de predicción es, en este caso, el ARMA(47,44). Se requiere, por tanto, 91 parámetros para ajustar el modelo ARMA óptimo, lo cual es coherente con la mayor variabilidad que posee esta serie respecto a los casos anteriores.

Puede observarse en la Fig. 5.15 sobre un segmento de 0.4 días de la serie, que la interpolación lineal produce más efectos espúreos que para el modelo de altas frecuencias debido a la mayor variabilidad de la señal en el rango afectado. En cambio, incluso en esta situación puede apreciarse como el método ARMA reconstruye la señal sin introducir efectos espúreos. Los periodogramas (Fig. 5.16) se han representado esta vez en escala lineal para poder apreciar mejor el espaciamiento entre frecuencias.

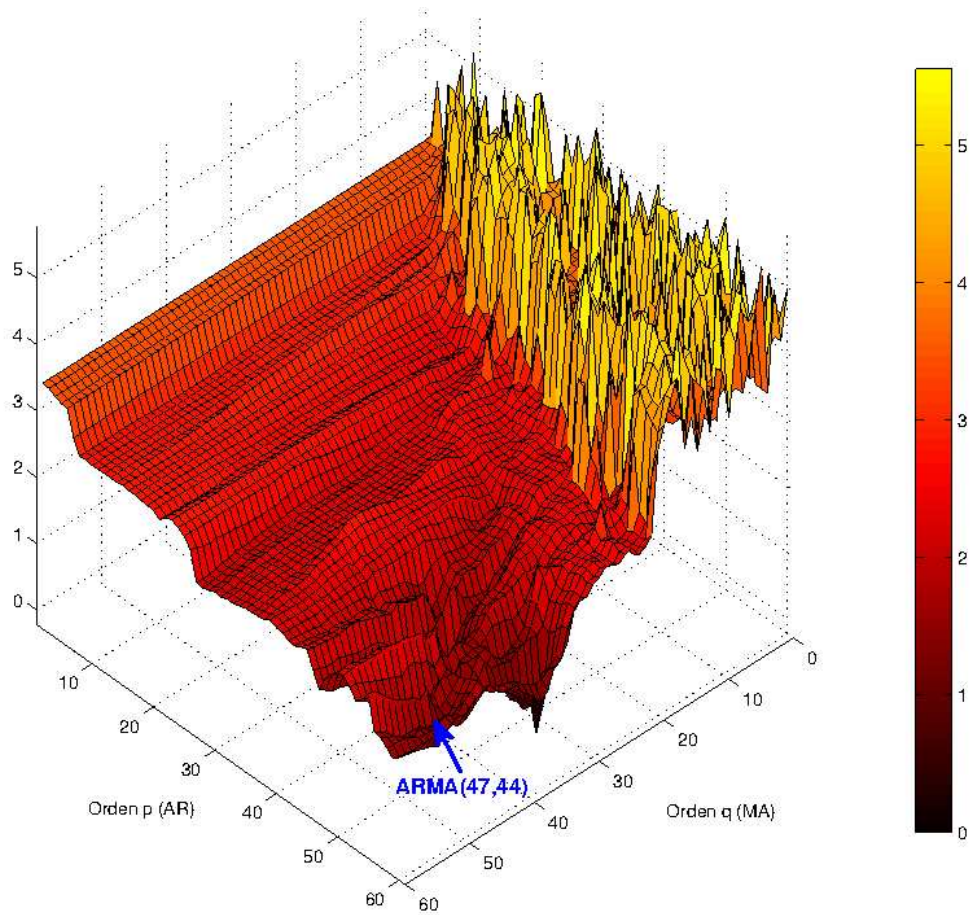


Figura 5.14: Estimación del orden ARMA(p,q) para un modelo de señal con 31 componentes armónicas y ruido blanco gaussiano. Nótese que el modelo óptimo es el ARMA(47,44).

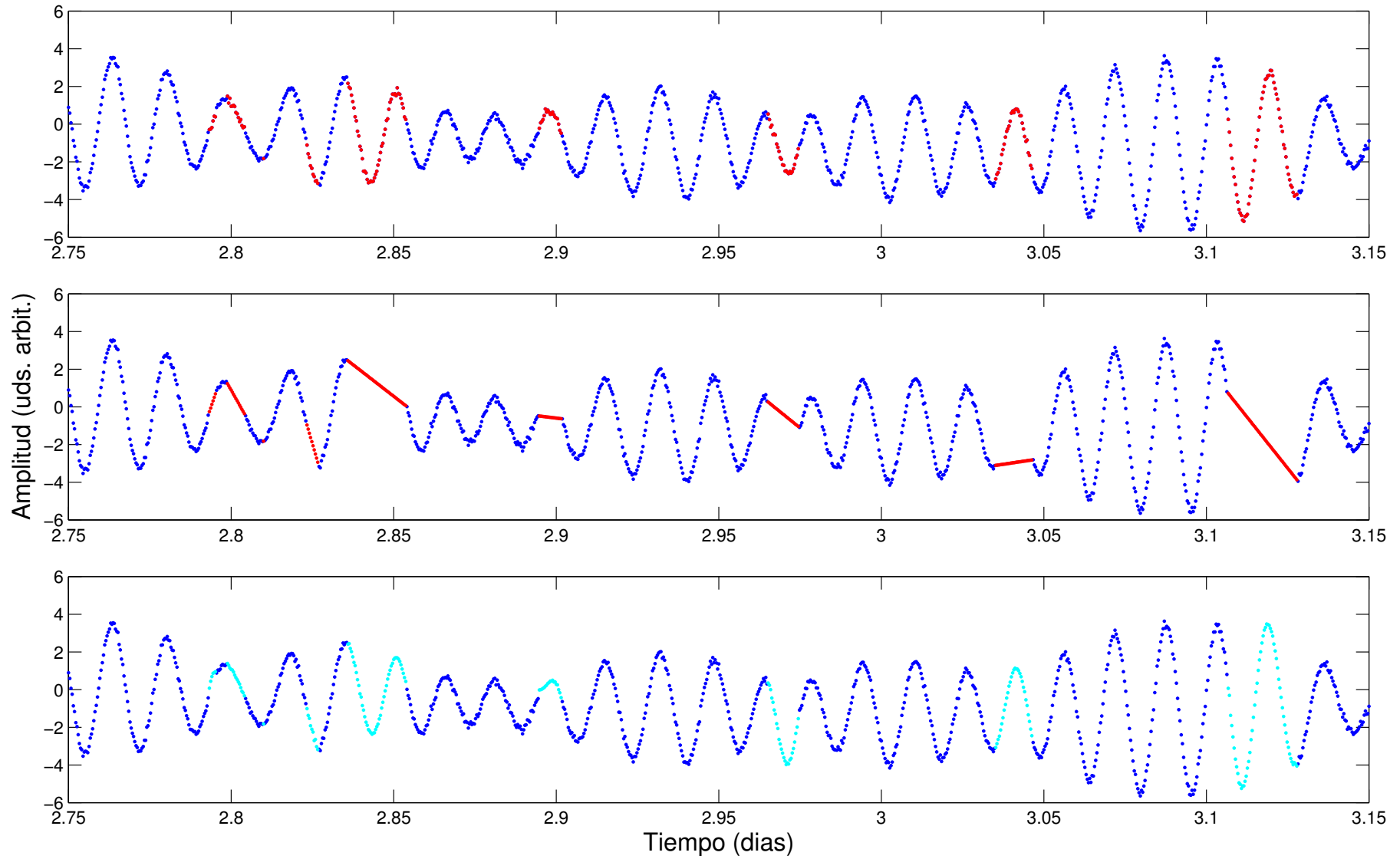


Figura 5.15: Comparación entre el método ARMA para el relleno de huecos e interpolación lineal para la serie del modelo con 31 frecuencias. En el panel superior se muestra un segmento de 0.4 días y en el panel central e inferior el mismo tras el relleno de huecos de los segmentos marcados en rojo y en cyan con el método de interpolación lineal y el método ARMA respectivamente.

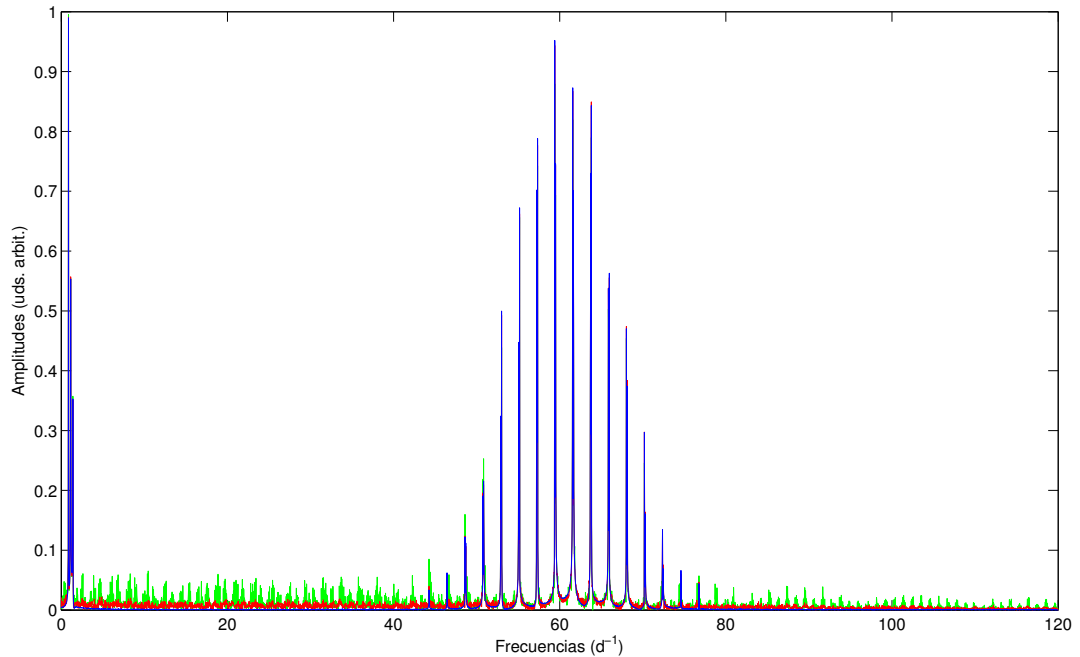


Figura 5.16: Periodograma de la serie sin huecos del modelo de 31 frecuencias (en azul) comparada con el periodograma de la serie con huecos interpolados con ARMA (en rojo) y linealmente (en verde).

Nótese que, si bien los picos de las frecuencias principales aparecen diferenciados con ambos métodos, en el caso de la interpolación lineal aparecen numerosas frecuencias espurias que, como demostraremos a continuación, tienen amplitudes suficientemente altas como para considerarse significativas y pueden ser confundidas con las componentes armónicas de la señal. Identificamos las frecuencias significativas en cada uno de los espectros a partir del algoritmo FRECS basado en una FFT (ver Apéndice A). Mediante este algoritmo se ha detectado las 31 frecuencias de la serie original en el espectro de la serie interpolada mediante ARMA (ver tabla 5.5 y 5.6), en cambio, en el espectro de la serie interpolada linealmente se detectan 34 frecuencias de las cuales 6 son frecuencias espurias producidas por la ventana espectral. Por otro lado, las componentes de $537.5 \mu\text{Hz}$, $863.5 \mu\text{Hz}$ y $888.5 \mu\text{Hz}$ no se han podido detectar en este espectro.

Con la lista de frecuencias de cada uno de los 4 espectros hallados (serie original, serie interpolada mediante ARMA, interpolada linealmente, y con huecos), hacemos el ajuste no-lineal de las frecuencias para obtener las amplitudes y las fases correspondientes. Estos ajustes nos dan los resultados mostrados en la tabla 5.5 y 5.6. Las amplitudes halladas a partir de la serie interpolada con ARMA difieren un 2.25 % con las originales con un máximo en torno al 7 % para la frecuencia de $837.5 \mu\text{Hz}$, sin que haya una variación significativa con el rango de frecuencias estudiado. En el caso de la serie interpolada linealmente las amplitudes difieren un 12.22 % en promedio

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

y llegando a variaciones de más del 100 % (para la frecuencia de $512.5 \mu\text{Hz}$). Estos resultados son similares a los que se obtienen a partir de la serie con huecos, en cuyo caso se detectan más frecuencias espurias pero la variación media de las amplitudes de las componentes de la señal detectadas está en torno a un 15 %.

Respecto a las fases, encontramos diferencias como máximo de 0.43 rads y 0.13 rads en promedio entre las fases ajustadas a partir de la serie interpolada mediante ARMA y las fases ajustadas a partir de la serie original. Las diferencias con las fases ajustadas a partir de la serie interpolada linealmente son de 0.15 rads en promedio pero llegan hasta $\sim \pi/4$ para la componente de $512.6 \mu\text{Hz}$. Un error de esa magnitud en la determinación de las fases da lugar a una inconsistencia en el ajuste de modelos teóricos de pulsación. Además, la determinación correcta de las amplitudes no es sólo importante para la detección de frecuencias por encima del nivel de ruido, sino que puede servir para determinar parámetros de las estrellas como el ángulo de inclinación. En efecto, se ha demostrado que para estrellas con ángulo de inclinación $i > 30$ el cociente de amplitudes de las componentes $m = \pm 1, 0$ de las oscilaciones de tipo dipolar es una función muy dependiente de i , con lo cual es posible determinar i a partir del análisis de frecuencias (Gizon and Solanki, 2003). Todo esto subraya la importancia de utilizar un método de interpolación fiable.

Por último, estudiamos los patrones en el espectro de potencias de la serie original sin huecos y de las 2 series interpoladas: linealmente y con ARMA. Para ello usamos el código desarrollado por García Hernández (2011) que se basa en una exploración de las periodicidades en la lista de frecuencias detectadas mediante el espectro de potencias de un tren de deltas. Las periodicidades así obtenidas son la mitad del espaciamiento en frecuencias. En la Fig. 5.17 podemos observar como es posible hallar las 2 periodicidades de la serie original (de $1 \mu\text{Hz}$ y $50 \mu\text{Hz}$) a partir de la serie interpolada mediante ARMA. En cambio, no es posible detectar la pequeña separación ($1 \mu\text{Hz}$) a partir de la serie interpolada linealmente. Además, el pico principal correspondiente al espaciamiento en frecuencias de $50 \mu\text{Hz}$ pierde amplitud cuando se interpola linealmente la serie.

Esta simulación numérica representa un caso ideal que en la práctica no se encuentra ya que las periodicidades entre frecuencias en el espectro de las estrellas δ Sct no son exactas sino que son cuasiperiodicidades. Sin embargo, con este sencillo ejemplo demostramos que es posible mejorar la detección de cuasiperiodicidades en el espectro de las δ Sct.

5.1 Tests del método mediante simulaciones numéricas

n	f_n^0 (μHz)	f_n^{ar} (μHz)	f_n^{gap} (μHz)	f_n^{lin} (μHz)	A_n^0	A_n^{ar}	A_n^{gap}	A_n^{lin}	φ_n^0	φ_n^{ar}	φ_n^{gap}	φ_n^{lin}
1	9.992	9.992	10.012	10.012	0.999	0.990	0.912	1.004	1.772	1.770	1.595	1.603
2	12.982	12.982	13.002	12.982	0.660	0.665	0.606	0.664	0.920	0.919	0.739	0.926
3	16.127	16.127	16.127	16.127	0.401	0.398	0.357	0.399	3.609	3.616	3.600	3.601
4	512.500	512.500	183.295	29.032	0.044	0.047	0.069	0.065	3.907	3.771	1.573	3.750
5	537.500	537.500	364.043	44.946	0.069	0.066	0.076	0.058	2.708	2.765	4.721	1.938
6	562.519	562.577	489.082	121.065	0.103	0.103	0.060	0.069	5.132	4.703	6.255	2.763
7	563.484	563.465	512.404	440.162	0.151	0.158	0.091	0.060	3.485	3.644	4.747	5.606
8	587.500	587.481	514.159	489.120	0.213	0.217	0.082	0.062	2.832	2.983	1.586	5.953
9	588.503	588.503	515.162	512.404	0.290	0.296	0.072	0.089	5.292	5.276	4.079	4.697
10	612.519	612.558	562.481	514.178	0.378	0.361	0.147	0.069	0.467	0.154	5.179	1.294
11	613.484	613.484	563.503	562.500	0.482	0.478	0.146	0.144	4.887	4.894	3.763	5.077
12	637.500	637.481	587.481	563.503	0.595	0.581	0.250	0.150	2.207	2.345	2.909	3.683
13	638.503	638.522	588.503	587.481	0.706	0.691	0.298	0.255	4.490	4.341	5.361	2.890
14	662.481	662.461	612.519	588.503	0.804	0.777	0.369	0.309	5.096	5.242	0.371	5.337
15	663.503	663.522	613.484	612.539	0.890	0.867	0.414	0.374	0.078	6.209	4.867	0.237
16	687.500	687.500	637.481	613.465	0.962	0.955	0.563	0.420	0.944	0.942	2.333	5.027
17	688.503	688.503	638.503	637.481	0.995	0.985	0.660	0.572	3.554	3.559	4.505	2.336
18	712.500	712.481	662.481	638.503	0.995	0.991	0.705	0.683	5.460	5.633	5.089	4.510

Tabla 5.5: Frecuencias estimadas por medio de un ajuste de mínimos cuadrados no-lineal a partir de la serie generada con el modelo de 31 frecuencias sin huecos f_n^0 , con interpolación ARMA f_n^{ar} , con huecos f_n^{gap} , y con interpolación lineal f_n^{lin} , y las amplitudes y fases correspondientes.

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

n	f_n^0 (μHz)	f_n^{ar} (μHz)	f_n^{gap} (μHz)	f_n^{lin} (μHz)	A_n^0	A_n^{ar}	A_n^{gap}	A_n^{lin}	φ_n^0	φ_n^{ar}	φ_n^{gap}	φ_n^{lin}
19	713.503	713.522	663.522	662.481	0.962	0.949	0.805	0.727	5.663	5.512	6.198	5.089
20	737.500	737.519	687.500	663.522	0.895	0.889	0.852	0.832	3.327	3.162	0.940	6.197
21	738.484	738.465	688.503	687.500	0.809	0.796	0.899	0.888	0.769	0.929	3.561	0.941
22	762.500	762.481	712.481	688.503	0.706	0.693	0.895	0.927	0.536	0.690	5.641	3.560
23	763.503	763.522	713.522	712.481	0.595	0.578	0.867	0.919	2.808	2.680	5.481	5.633
24	787.500	787.519	737.519	713.522	0.484	0.481	0.817	0.889	2.319	2.172	3.170	5.492
25	788.484	788.465	738.465	737.519	0.380	0.376	0.742	0.833	5.883	6.042	0.917	3.166
26	812.500	812.500	762.500	738.465	0.290	0.290	0.649	0.753	2.873	2.866	0.568	0.920
27	813.503	813.484	763.503	762.500	0.212	0.209	0.515	0.663	6.025	6.179	2.842	0.552
28	837.500	837.481	787.539	763.503	0.151	0.140	0.460	0.535	5.231	5.383	2.069	2.825
29	838.503	838.542	788.484	787.539	0.103	0.101	0.385	0.462	5.731	5.317	5.826	2.045
30	863.503	863.503	812.481	788.484	0.069	0.068	0.219	0.373	4.579	4.612	3.049	5.857
31	888.503	888.503	813.503	812.500	0.044	0.047	0.217	0.235	3.265	3.214	0.006	2.883
32	-	-	837.365	813.484	-	-	0.081	0.205	-	-	0.002	0.056
33	-	-	838.542	837.442	-	-	0.133	0.101	-	-	4.833	5.667
34	-	-	860.822	838.522	-	-	0.082	0.116	-	-	0.439	5.202
35	-	-	863.503	-	-	-	0.069	-	-	-	5.463	-
36	-	-	936.902	-	-	-	0.062	-	-	-	1.646	-
37	-	-	985.860	-	-	-	0.067	-	-	-	1.908	-

Tabla 5.6: Continuación de la tabla 5.5

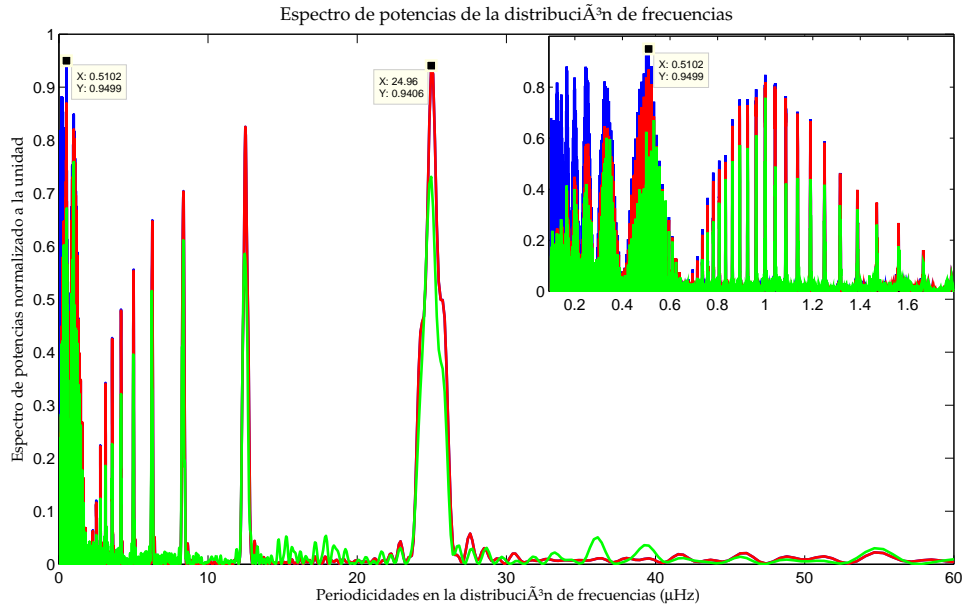


Figura 5.17: Periodicidades en el rango de 0.1 a 60 μHz de la distribución de frecuencias detectadas en los periodograma de la serie sin huecos (en azul) comparada con las periodicidades del periodograma de la serie con huecos interpolados con ARMA (en rojo) y linealmente (en verde). En el inserto se representa la región de bajas frecuencias donde se haya la pequeña separación $\delta\nu$. Los valores reales se calculan como el doble de los hallados en la figura.

5.2 Curvas de luz de CoRoT

El paso de CoRoT por la Anomalía Suratlántica introduce la mayoría de las perturbaciones presentes en los datos del satélite, como fue mencionado anteriormente. Estas perturbaciones sobre el satélite provocan que la cobertura de datos no sea del 100%. Hay aproximadamente (depende del dataset) un 10% de datos no-válidos en cada curva de luz observada por el satélite que han sido corregidos por el pipeline de procesado de datos desde Tierra mediante interpolaciones lineales. En esta sección estudiamos los efectos de la interpolación lineal y como estos se minimizan al utilizar interpolación ARMA. Analizaremos tres curvas de luz observadas procedentes de estrellas de distintos tipos de variabilidad: tipo solar, δ Sct, y Be.

5.2.1 Estrella tipo-solar: HD 49933

Las oscilaciones que ocurren en las estrellas de tipo solar se caracterizan porque el mecanismo de excitación es estocástico y además se ven amortiguadas por la convección. Esto se traduce en un espectro ruidoso y un ensanchamiento de las líneas de las frecuencias debido al amortiguamiento. Además, las cavidades internas de la estrella dan lugar al mecanismo de resonancia causante del exceso de potencia en

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

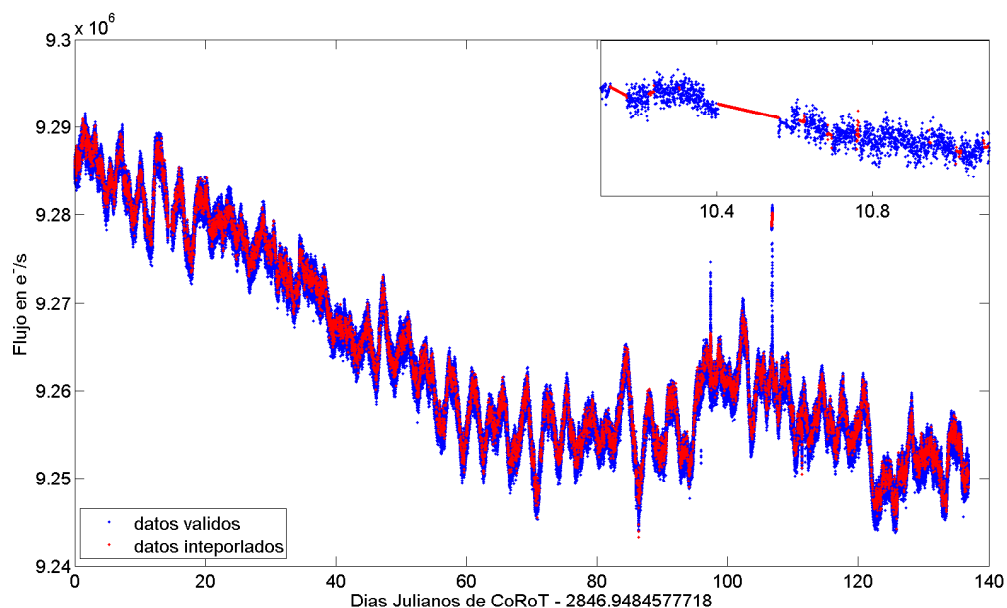


Figura 5.18: Curva de luz de la estrella de tipo solar HD 49933. Los puntos marcados en rojo se corresponden con los datos que han sido interpolados linealmente en el pipeline de los datos de CoRoT de nivel 2. En el inserto se representa un segmento de un día.

forma de gaussiana en un conjunto de frecuencias equiespaciadas que es la característica principal del espectro de este tipo de estrellas. Las frecuencias excitadas suelen corresponder a periodos del orden de minutos.

HD 49933 es una estrella F5V de la secuencia principal que ha sido observada durante el Initial Run (60 días) y el primer Long Run (137 días) en el anticentro galáctico (LRa01). En este caso estudiamos el efecto de la interpolación que sustituye los conjuntos de datos no-válidos de la serie temporal del LRa01, que puede observarse en la Fig. 5.18.

En la figura puede observarse una tendencia a decaer del flujo debido al envejecimiento de la CCD, y otras variaciones de largo periodo debido, probablemente, a la actividad estelar. Antes de estudiar los efectos de la interpolación filtramos estas variaciones no periódicas de larga duración restando a la curva de luz un polinomio de tercer grado. Para que el proceso de relleno de huecos funcione correctamente es necesario corregir estas variaciones así como saltos y otros efectos instrumentales que puedan dar lugar a una mala elección del modelo de la serie.

El orden óptimo del modelo ARMA para el segmento más largo sin datos inválidos, de 640 puntos, es el (37,0), es decir, un modelo puramente autoregresivo con 37 términos. Este modelo se ha usado para interpolar los segmentos de datos no-válidos contenidos en los 369601 puntos de la serie de datos de nivel 2 de HD 49933. En la Fig. 5.19 se muestran detalles de la interpolación ARMA comparada con la interpo-

lación lineal. Pese a que, como se menciona en Appourchaux et al. (2008), el tamaño de los huecos es de 9 minutos, los huecos más grandes tienen un tamaño en torno a 0.2 días o 168 minutos y en ellos se pierde señal de un amplio rango de frecuencias. Esta serie tiene, además, un nivel de ruido muy alto con lo cual la interpolación lineal es bastante errática como puede apreciarse en el segundo panel. No obstante, como se puede ver, por ejemplo, en el segmento interpolado entre 1.4 y 1.6 días, la interpolación ARMA da lugar a una estructura fina de apariencia similar al ruido. Esto arroja dudas sobre la naturaleza de esas variaciones, ya que el algoritmo sólo extrapola la señal y no el ruido en los huecos. En la Fig. 5.20 no es posible apreciar gran diferencia entre los periodogramas de la serie interpolada linealmente y la serie interpolada mediante ARMA debido a la gran densidad de picos. Esto se debe al alto nivel de ruido, al gran número de frecuencias excitadas de este tipo de estrellas y, por otro lado, a que las líneas de las frecuencias, al ser el mecanismo de excitación estocástico como se dijo más arriba, se ensanchan dando lugar a una lorentziana en lugar de una línea. Por ello, mostramos un inserto en la figura en el que puede verse el rango en el que en los estudios de esta estrella se detectaron los modos-p. En el inserto los picos correspondientes al periodograma de la serie interpolada mediante ARMA tienen una mayor amplitud que los que corresponden a la interpolación lineal.

Aunque en Appourchaux et al. (2008) y en estudios posteriores basados en este (Benomar et al., 2009), se hace uso de la interpolación lineal, en la Fig. 5.20 hemos visto que el impacto de la interpolación en los huecos no es despreciable para la detección y determinación de los parámetros de las líneas.

5.2.2 δ Scuti: HD 174966

HD 174966 es una estrella pulsante de tipo δ Sct de masa intermedia que se halla en la secuencia principal. Las δ Sct pertenecen al grupo de estrellas conocidas como pulsantes clásicas porque se encuentran en el diagrama HR en la banda de inestabilidad clásica, donde también se hallan las estrellas Cefeidas. Las estrellas de la banda de inestabilidad clásica se vienen estudiando teóricamente desde hace décadas y por ello poder interpretar las observaciones es de sumo interés. Además, las observaciones durante la última década por Garrido and Poretti (2004) y Breger et al. (2005) han mostrado que el número de frecuencias detectadas en estas estrellas es mayor de lo que se creía y este aumenta conforme mejora la calidad de las observaciones.

Hasta años recientes no existían estudios sobre la variabilidad de HD 174966 pero con el lanzamiento de CoRoT se ha podido estudiar sus pulsaciones (García Hernández et al., 2013). En el estudio mencionado se ha confirmado la detección de un patrón similar a la gran separación $\Delta\nu$. Esta periodicidad se ha confirmado (Suárez et al., 2014) se aproxima al régimen asintótico de Tassoul y que simulamos con un modelo

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

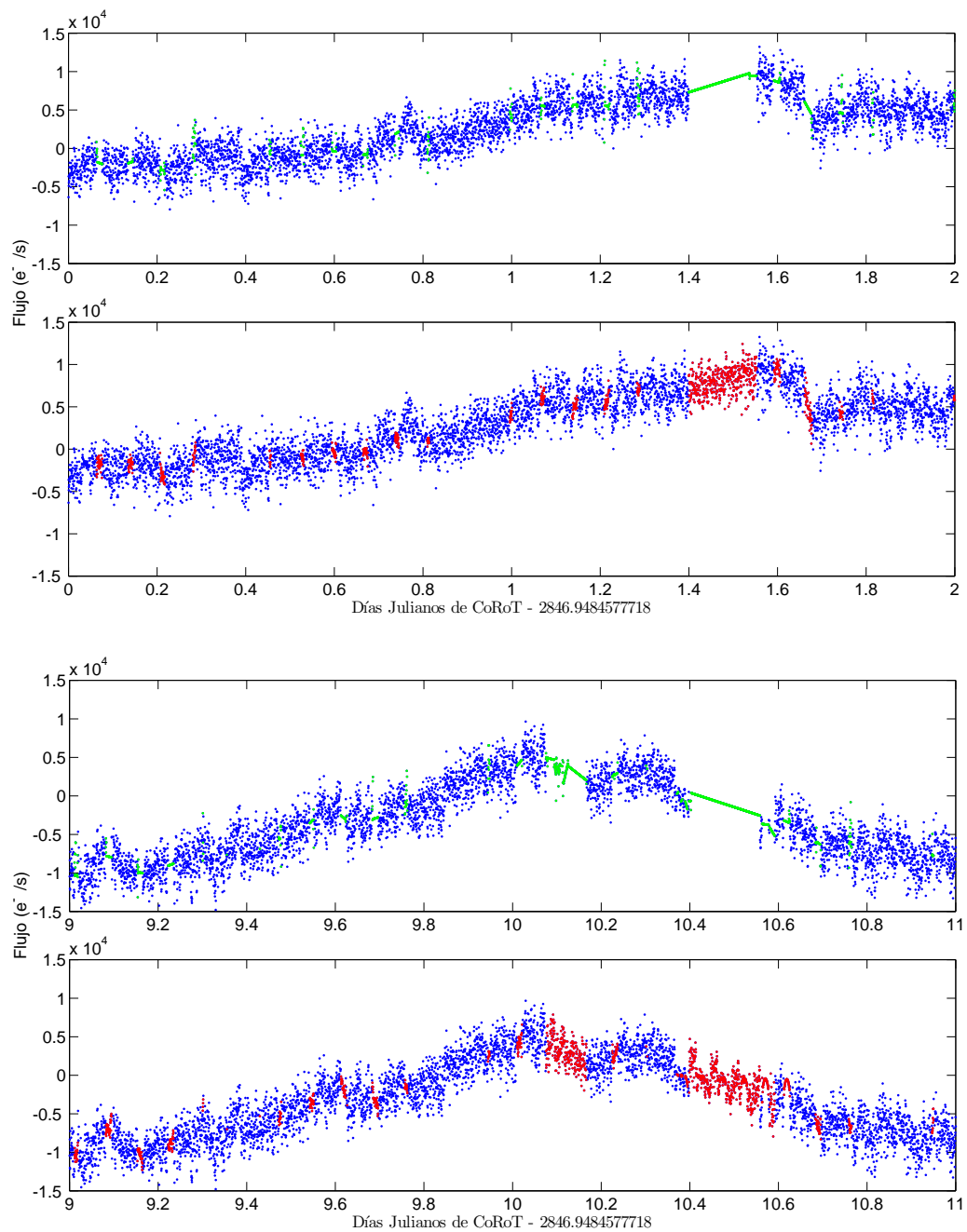


Figura 5.19: Comparación entre el método ARMA para el relleno de huecos (en rojo) e interpolación lineal (en verde) para la curva de luz de la estrella HD 49933 observada por CoRoT. En los paneles se muestran los huecos más grandes de la serie. Cada panel muestra un segmento distinto de 2 días de duración.

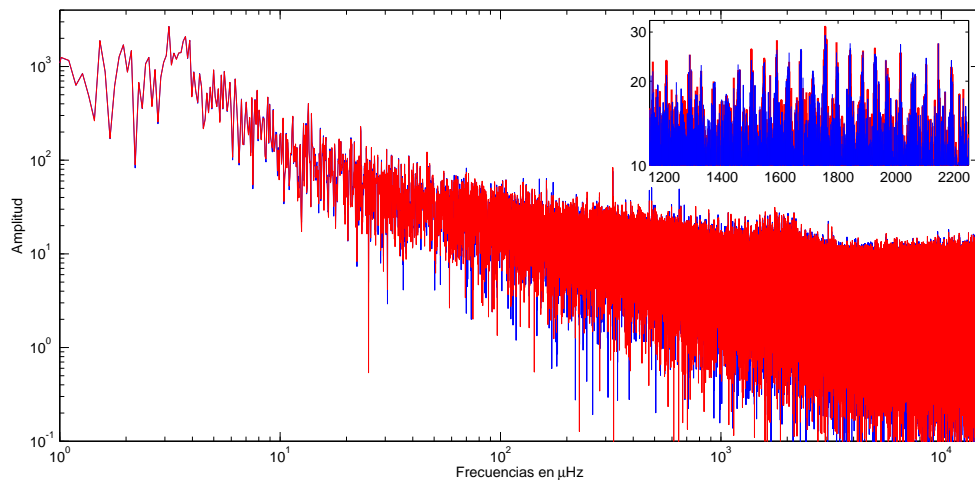


Figura 5.20: Periodograma en escala logarítmica de la curva de luz de la estrella HD 49933 interpolada linealmente en los huecos (en azul) comparada con el periodograma de la serie con huecos interpolados con ARMA (en rojo). En el inset se muestra el rango de 1100 a 2300 μHz donde se encuentran los modos-p solares. Nótese que las amplitudes de espectro interpolado mediante ARMA son mayores que las del espectro interpolado linealmente.

analítico en el último test de la sección anterior. Utilizando herramientas de análisis de series temporales como PERIOD04 (Lenz and Breger, 2005) y SIGSPEC (Reegen, 2007) para la detección de frecuencias sobre la serie con huecos encuentran 185 frecuencias significativas en el rango de 0 a 900 μHz . En la Fig. 5.21 se muestra un segmento de la curva de luz donde puede observarse el efecto de las interpolaciones lineales, marcadas en rojo en la figura. En este caso el nivel de ruido es mucho menor que en el caso estudiado en el apartado anterior y pueden apreciarse rasgos de la curva de luz como el batido entre frecuencias.

El orden óptimo para el modelo ARMA en este caso es el (34,6), es decir, un modelo con 40 términos de los cuáles 6 son de media móvil y 34 son autoregresivos. Este modelo se ha usado para interpolar los segmentos de datos no-válidos contenidos en los 73433 puntos de la serie de datos de nivel 2 de HD 174966. En la Fig. 5.22 se muestran detalles de la interpolación ARMA comparada con la interpolación lineal. Puede apreciarse como algunas de las interpolaciones (véase p.ej. el rango de 1.38 a 1.40 días) lineales en las que se destruye la señal se han corregido en el panel inferior donde mostramos la interpolación ARMA. Esto es determinante para la correcta determinación de las fases y también para identificar las frecuencias que componen la señal y poder estudiar la cuasiperiodicidad.

Si bien, visualmente, en este caso el efecto de la interpolación parece ser menor en la curva de luz que en el caso anterior, en la Fig. 5.23 puede apreciarse como la interpolación ARMA elimina toda la contribución de la ventana espectral que la interpolación

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

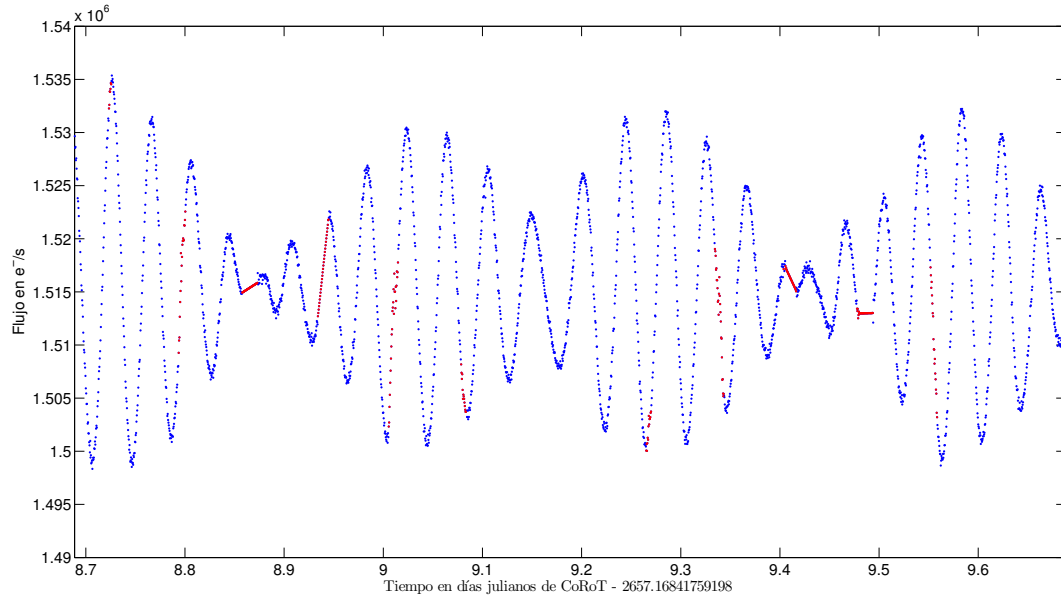


Figura 5.21: Curva de luz de la estrella δ Sct HD 174966. Los puntos marcados en rojo se corresponden con los datos que han sido interpolados linealmente por el pipeline que procesa los datos de CoRoT de nivel 2.

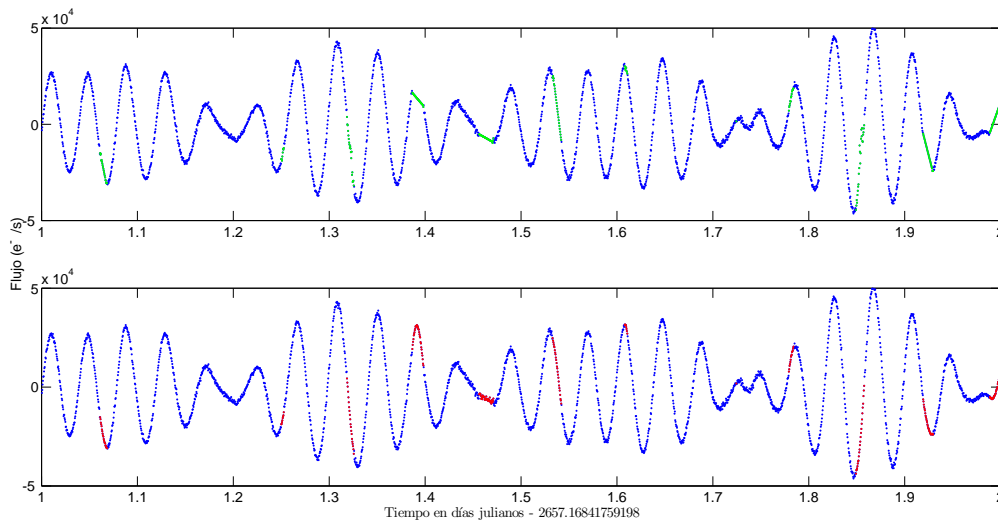


Figura 5.22: Comparación entre el método ARMA para el relleno de huecos (en rojo) e interpolación lineal (en verde) para la curva de luz de la estrella HD 174966 observada por CoRoT.

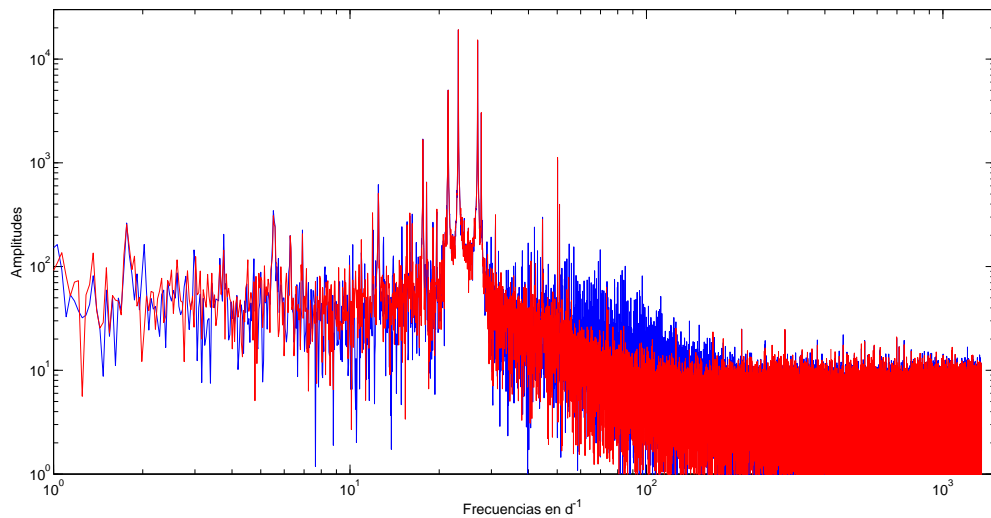


Figura 5.23: Periodograma en escala logarítmica de la curva de luz de la estrella HD 174966 interpolada linealmente en los huecos (en azul) comparada con el periodograma de la serie con huecos interpolados con ARMA (en rojo). Nótese como el efecto de la ventana espectral se mantiene con la interpolación lineal.

lineal (en azul en la figura) no es capaz de eliminar. Esto repercute directamente en el estudio de las cuasiperiodicidades como puede apreciarse en la Fig. 5.24. Para obtener la lista de frecuencias de cada una de las 2 series (interpolada linealmente e interpolada con ARMA) se ha utilizado la rutina FRECS (ver apéndice A). Posteriormente, haciendo uso de la técnica para el estudio de las periodicidades desarrollada por García Hernández [García Hernández \(2011\)](#), se han obtenido las gráficas mostradas en la Fig. 5.24 para diferentes sets de frecuencias. Los resultados son coherentes con los que se muestran en [García Hernández et al. \(2013\)](#), donde se estudian las periodicidades para esta misma estrella pero, en efecto, puede verse que se gana en amplitud al usar el set de frecuencias detectadas a partir de la serie interpolada mediante ARMA respecto a la serie interpolada linealmente. Esto se debe a una mejor determinación de los valores de las frecuencias y a que el conjunto de frecuencias obtenidas mediante la serie interpolada linealmente contiene frecuencias espurias.

En conclusión se deduce que el método de interpolación es crucial para el estudio de las periodicidades en este tipo de estrellas. Por otra parte, en los segmentos interpolados mediante ARMA no encontramos la estructura fina con apariencia de ruido que encontramos en el caso anterior. Los segmentos interpolados mantienen las propiedades a una escala reducida de los segmentos de datos originales. Se descarta, por tanto, que los efectos mencionados sean metodológicos.

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

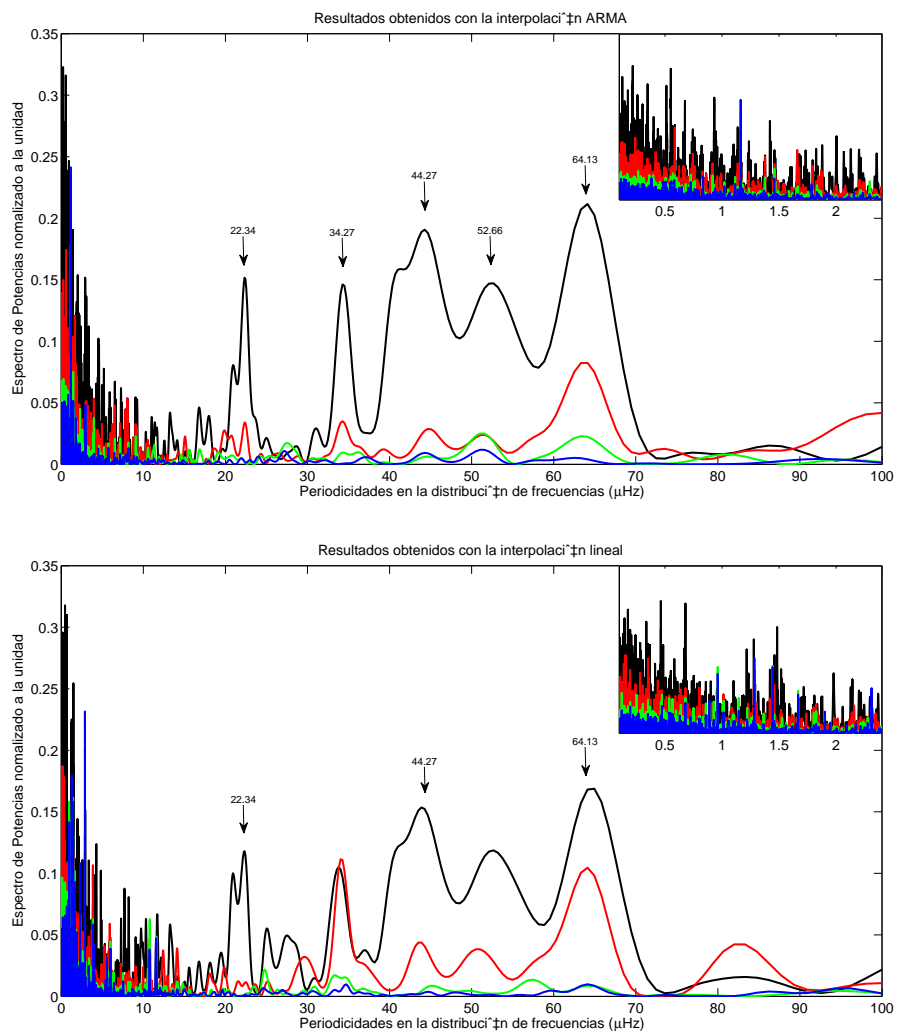


Figura 5.24: Periodicidades de HD 174966 en el rango de 0.1 a 100 μHz de la distribución de frecuencias detectadas en los periodograma de la serie interpolada con ARMA (panel superior) comparada con las periodicidades del periodograma de la serie con huecos interpolados linealmente (panel inferior). Las líneas negra, roja y verde se corresponden respectivamente con los subsets de 30, 60 y 114 frecuencias. En el inserto se representa solamente el rango de bajas frecuencias.

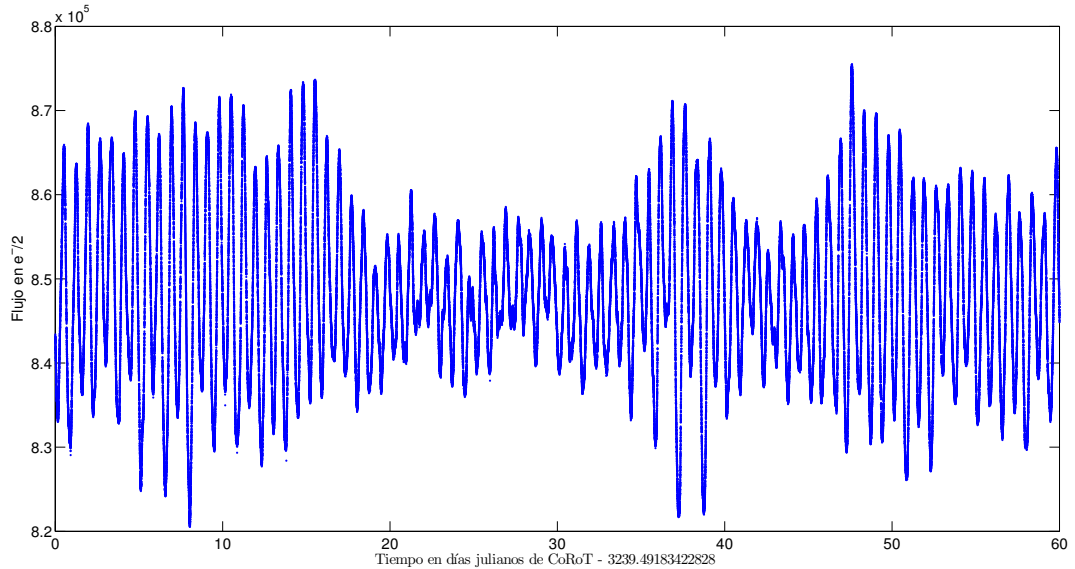


Figura 5.25: 60 días de la curva de luz de la estrella Be HD 51193 observada por CoRoT.

5.2.3 Estrella Be: HD 51193

La estrella HD 51193 (CoRoT 714) es una estrella de tipo espectral B1 con pulsaciones de tipo Be que ha sido observada durante el segundo Long Run hacia el anticentro galáctico (LRa02). Los datos que se muestran aquí no han sido publicados anteriormente pero esta estrella fue observada entre el 2004 y el 2006 desde el Observatorio de Sierra Nevada (OSN) y se encontró una frecuencia de 2.66 cd^{-1} en los filtros vby del fotómetro Stromgren (Gutiérrez-Soto et al., 2007).

Las estrellas Be son estrellas de tipo espectral B que muestran emisión en las líneas de Balmer. Esta emisión se produce por la presencia de un disco ecuatorial circunestelar que es alimentado por eventos en los que se produce una pérdida de masa estelar. Estas estrellas se caracterizan por tener dos tipos de variabilidad fotométrica: variabilidad de largo periodo (semanas o meses) y variabilidad de corto periodo (0.1 a 2 días). La primera se produce debido a las variaciones de densidad y tamaño del disco circunestelar, y la segunda se atribuye a pulsaciones no-radiales.

Para caracterizar su variabilidad y hacer uso de modelos que expliquen el origen de esta es necesario que los datos puedan ser analizados sin huecos para poder estudiar la evolución de las amplitudes y fases con el tiempo.

La serie completa de HD 51193 contiene 114.41 días pero en la fig. 5.25 mostramos los primeros 60 días observados por CoRoT para visualizar las variaciones de la curva de luz. En esta figura puede apreciarse que la amplitud de las oscilaciones varían de forma bastante irregular. Esto es una característica típica de este tipo de estrellas pulsantes. El orden óptimo del modelo ARMA para esta serie según el criterio de

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

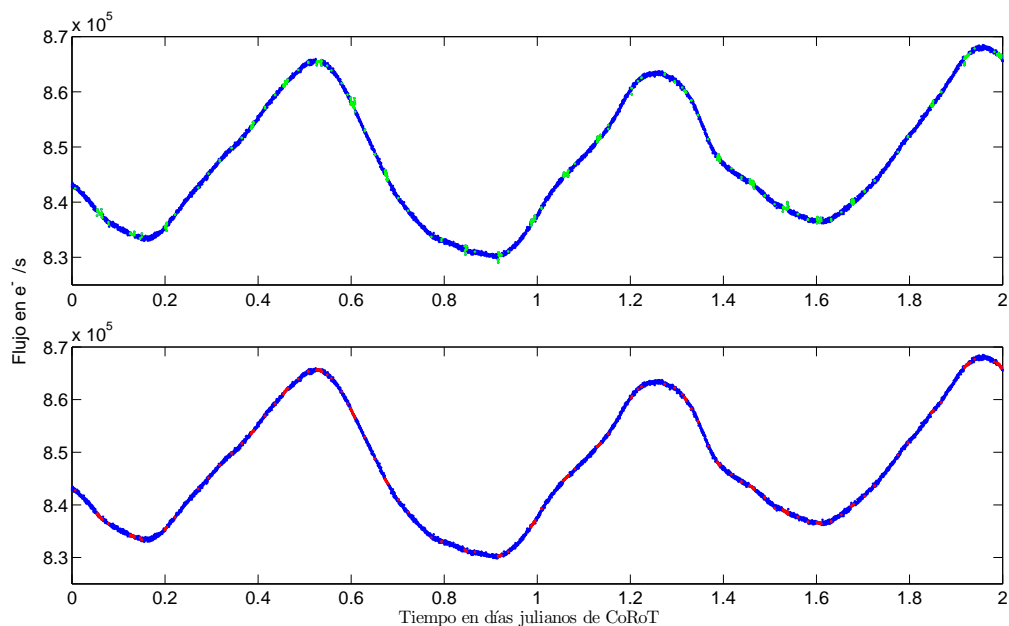


Figura 5.26: Comparación entre el método ARMA para el relleno de huecos (en rojo) e interpolación lineal (en verde) para la curva de luz de la estrella Be HD 51193 observada por CoRoT.

Akaike resulta ser el (24,17). El modelo que usamos para la interpolación de los huecos consta, por tanto, de 41 coeficientes de los cuales 17 son de media móvil y 24 son autoregresivos. Con este modelo se ha reconstruido el muestreo uniforme de la serie de 308917 puntos.

En la Fig.5.26 se muestran detalles de la interpolación ARMA comparada con la interpolación lineal. Los detalles que aparecen en esta figura no muestran una diferencia notable entre los resultados de ambas interpolaciones. Así, aparentemente, no sería necesario utilizar una interpolación más sofisticada que la interpolación lineal para una curva de luz con variaciones de largo periodo como es esta. Esto sería coherente con los resultados obtenidos en el test numérico de bajas frecuencias. No obstante, cuando estudiamos el periodograma correspondiente a cada serie los resultados que encontramos son bien distintos.

Como puede verse en la Fig. 5.27 los alias que aparecen en el periodograma de la serie interpolada linealmente han desaparecido casi por completo en la serie interpolada mediante ARMA. Solamente aparece un pico en la frecuencia orbital (13.97 cd^{-1}) debido a que no es posible eliminar la modulación del satélite con la interpolación. Debido a la sensibilidad de los detectores del satélite a las variaciones de temperatura la función respuesta del satélite es función de su órbita también con lo que se produce una modulación. Este efecto es mucho más marcado en la cámara de exoplanetas pero aparece también en algunos objetivos del seismocampo como este.

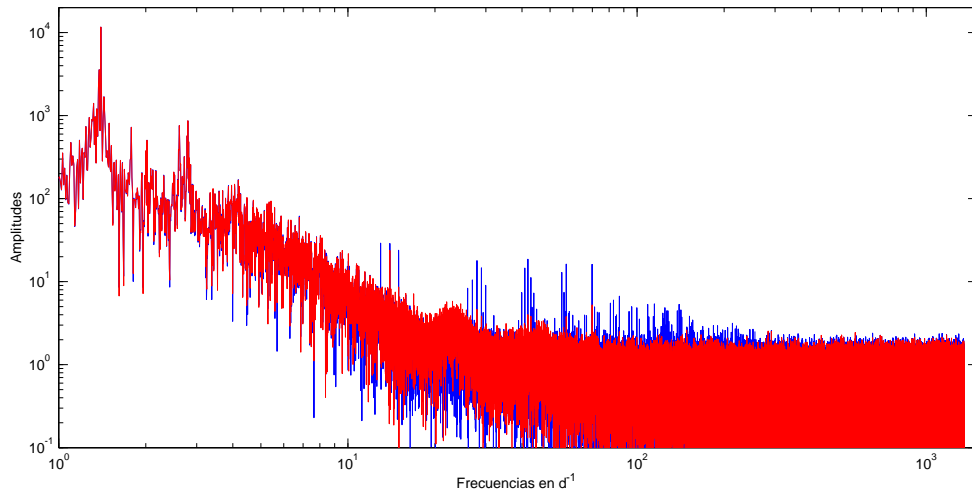


Figura 5.27: Periodograma en escala logarítmica de la curva de luz de la estrella HD 51193 interpolada linealmente en los huecos (en azul) comparada con el periodograma de la serie con huecos interpolados con ARMA (en rojo). Nótese como el efecto de la ventana espectral se mantiene con la interpolación lineal.

Por otro lado, el hecho de que el espectro de una estrella Be con pulsaciones de bajas frecuencias se vea mucho más afectado por los alias que en el test numérico llevado a cabo en § 5.1.1.1 hace pensar que la señal que se haya bajo esta curva de luz es una función de distinto tipo a la que corresponde al modelo usado en el test, es decir, no se comporta como una función analítica.

5. RESULTADOS: SERIES SINTÉTICAS Y CURVAS DE LUZ DE COROT

Parte III

Estudio de la validez del desarrollo en serie de Fourier de una serie temporal

Gracias a la calidad sin precedentes de los datos obtenidos por misiones espaciales como CoRoT o *Kepler* se ha conseguido mejorar los modelos teóricos de las pulsaciones estelares y con ello realizar una mejor inferencia de la estructura estelar. Pero para entender correctamente la naturaleza de los procesos observados, el aparato matemático usado para interpretar los datos debe poder reproducir la estructura entera de los datos de forma numérica, dejando sólo una serie aleatoria independiente como residuos. En la práctica esto raramente ocurre y los residuos, por pequeños que sean, preservan cierta estructura que hace que sean muy distintos de una secuencia aleatoria independiente y en muchos casos resultan tener una distribución no gaussiana. Esto es especialmente evidente en algunas estrellas pulsantes δ Sct (Poretti et al. 2009; García Hernández et al. 2009) y γ Dor (Chapellier et al., 2011) observadas por CoRoT.

Desde que se realizaron los primeros análisis el aparato matemático usado en astrosismología se ha basado series de Fourier (i.e. periodograma Lomb-Scargle, Scargle 1982). El análisis de Fourier es una técnica útil para la detección de frecuencias pero tiene una variedad de dificultades que no se consideran siempre en detalle en la literatura. Si la secuencia analizada pertenece a un conjunto de datos no-periódico o cuasiperiódico, el número de componentes espectrales necesitado para representar la señal puede ser tan grande que la información física proporcionada por su representación de Fourier pierda su utilidad. Restringir el número de componentes espectrales usado para el análisis armónico sólo incrementa la probabilidad de obtener información engañosa.

Por otro lado, para poder utilizar el potencial de las técnicas de Fourier el Teor. de Parseval ha de cumplirse. En su versión discreta este teorema dice que la suma del cuadrado del módulo de una secuencia es igual a la suma del cuadrado del módulo de su transformada de Fourier. Esto hace posible la representación espectral de la secuencia siempre que esta sea de cuadrado sumable. Esto es algo que se utiliza como hipótesis de partida pero nunca se verifica.

En esta parte de la memoria examinamos la correspondencia de algunas propiedades derivadas a partir de las componentes espectrales halladas en la curva de luz de la estrella HD 174936 con las técnicas clásicas de Fourier, con las propiedades que se pueden derivar directamente de los datos originales. Concretamente, examinamos una propiedad de estructura fina¹ que llamamos conectividad (Pascual-Granado et al., 2014) que está intrínsecamente relacionada con la analiticidad de la función continua que subyace a las series temporales.

¹Llamamos aquí estructura fina a variaciones intrínsecas de la serie a una escala del orden del muestreo Δt de la serie y que no se deben a ningún fenómeno estocástico.

De este modo, estamos evaluando la identidad entre la representación espectral obtenida y los datos originales, y al mismo tiempo evaluamos la validez de la hipótesis de integrabilidad del cuadrado de la señal.

El primer capítulo de esta parte está dedicado a sentar las bases matemáticas que sustentan las técnicas y los resultados que se presentan en los 2 capítulos siguientes. En el Cap. 7 definimos el concepto de conectividad, que es una función que describe de que forma se conectan los puntos de la función muestreada. Definida de este modo, la conectividad da una aproximación adecuada al concepto de diferenciabilidad para series temporales siempre y cuando la función sea de banda limitada y el muestreo sea suficiente. Usaremos 2 aproximaciones para hallar las conectividades, una está basada en los polinomios splines y la otra se basa en modelos autoregresivos. Los resultados de las pruebas que llevamos a cabo se presentan en el Cap. 8. El primer caso estudiado es la estrella HD 174936 para la cual hemos tratado de replicar los resultados obtenidos con los datos de CoRoT mediante una simulación numérica basada en el contenido en frecuencias para esta estrella publicado en la literatura. Además, para descartar que los efectos instrumentales puedan ser una causa de los fenómenos presentados, hicimos el test de analiticidad también en otras estrellas observadas mediante *Kepler*, y el Sol mediante SoHO.

Consideraciones sobre la consistencia del análisis de frecuencias

“It is a capital mistake to theorise before one has data. Insensibly one begins to twist facts to suit theories instead of theories to suit facts.”

The Adventures of Sherlock Holmes - A. C. Doyle

En general, la hipótesis más elemental que se puede hacer sobre las fluctuaciones de cualquier fenómeno es que éstas se distribuyen siguiendo una gaussiana ya que es la distribución con mayor entropía de entre todas las distribuciones de los reales posibles con una media y varianza determinadas. Además el Teorema Central del Límite ([Spiegel, 1992](#)) nos dice que la suma de N valores aleatorios independientes (siendo N un número entero suficientemente grande) tomados de una distribución con media y varianza definidas, da como resultado valores de una distribución gaussiana. Es decir, independientemente de qué distribución tengan esos valores, su suma seguirá una distribución gaussiana. Esto hace que la distribución gaussiana esté presente en la mayoría de los fenómenos físicos estudiados y se asume a menudo la validez de la hipótesis de gaussianidad (véase p.ej. la radiación cósmica de microondas ([Guth, 1981](#))). En § [6.1](#) discutimos la trascendencia de esta distribución.

Por otra parte, es tremendamente útil comparar con una distribución gaussiana ya que las desviaciones sobre ésta nos proporcionan información sobre el fenómeno físico que se estudia. Existe toda una plétora de mecanismos físicos distintos que pueden causar la aparición de desviaciones sobre la gaussiana en la distribución. Es, por tanto, muy complicado realizar de partida una descripción física de la no gaussianidad. No obstante, se puede inferir estudiando propiedades matemáticas de las series temporales que pueden inducir la no gaussianidad como son la no linealidad o la no-estacionaridad. En § [6.2](#) discutimos, además, cómo la no analiticidad de la función que subyace a las series temporales puede estar también en el origen de la no gaussianidad.

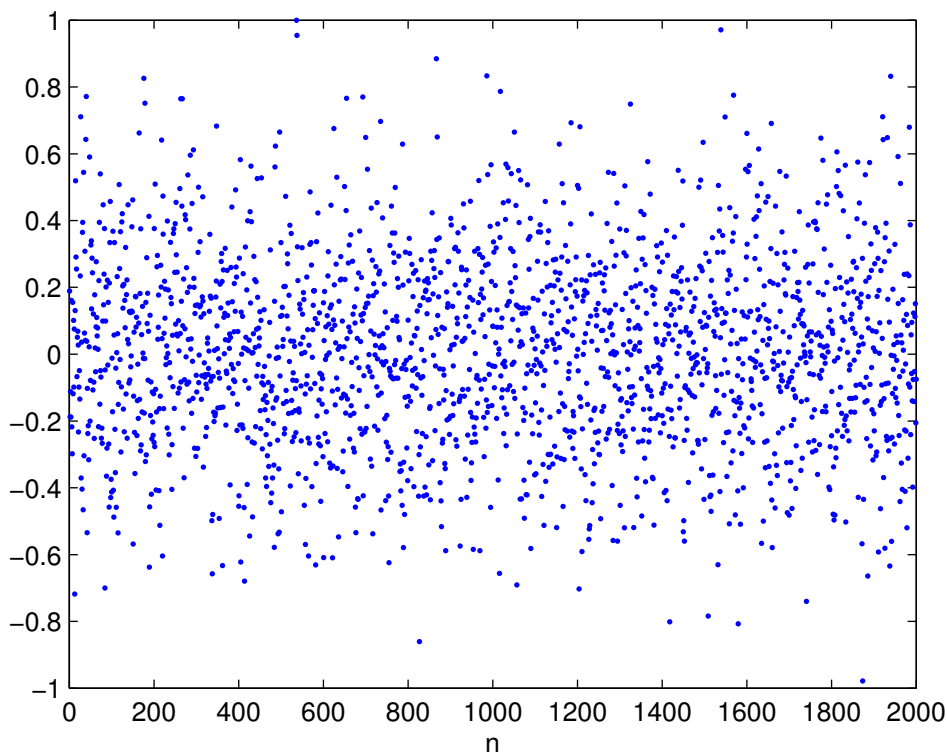


Figura 6.1: Serie de datos aleatorios generados según una distribución gaussiana de tipo $\mathcal{N}(0, 1)$ normalizada a la unidad. El valor de la serie numérica se representa en ordenadas y en abscisas el índice correspondiente. Nótese como los datos se distribuyen de manera suave en torno a la media.

La analiticidad, tal y como se definimos en § 7, es además condición necesaria para poder realizar un análisis armónico consistente. Esta condición se desarrolla en el § 6.3 de este capítulo.

6.1 La distribución gaussiana

Una variable X cuya densidad de probabilidad es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty \quad (6.1)$$

se dice que tiene una distribución gaussiana o normal con media μ , y varianza σ^2 . La notación usual para designar esa distribución es $\mathcal{N}(\mu, \sigma^2)$. La densidad de probabilidad de esta distribución tiene la bien conocida forma acampanada y las gráficas de las series que siguen esta distribución se caracterizan porque aparecen como un gradiente suave en torno a la media de la distribución ya que los saltos grandes en ordenadas son poco probables. En la Fig. 6.1 se ha representado una realización de una secuencia de datos que sigue esta distribución donde se puede apreciar estos detalles.

El rol preeminente que esta distribución juega en teoría de probabilidad se atribuye al Teor. Central del Límite. Este teorema se formula de la siguiente manera:

Sea $X_1, X_2, \dots, X_n, \dots$ una secuencia de variables aleatorias idénticamente distribuidas, cada una con una media finita μ y una varianza σ^2 , y S_n

$$S_n = \sum_{i=1}^n X_i \quad n = 1, 2, \dots$$

entonces cuando $n \rightarrow \infty$. la distribución de la secuencia

$$(S_n - n\mu)/(\sigma\sqrt{n}) \rightarrow \mathcal{N}(0, 1)$$

Y de manera equivalente se puede decir: $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$. Concretamente, es fácil probar para una distribución binomial o una distribución de Poisson que la suma de un número considerable de variables independientes da lugar a una distribución gaussiana.

En vista de este resultado es razonable suponer que los errores involucrados en las medidas de la mayoría de las cantidades físicas tengan distribuciones muy cercanas a la gaussiana dado que esos errores surgen usualmente del efecto combinado de un gran número de fuentes de error independientes. De este modo, la cantidad física medida como suma de un valor “verdadero” más un error tendría también una distribución aproximadamente gaussiana (Priestley, 1981).

Sin embargo, la trascendencia de esta distribución va más allá de su uso en inferencia estadística, ya que se ha de entender esta como la distribución más natural en cualquier fenómeno estocástico dado que maximiza la entropía. En efecto, dada la media μ y la varianza σ^2 de una distribución, la gaussiana es aquella que maximiza la función entropía de la información (estrechamente relacionada con la entropía termodinámica) que toma el valor $\ln(\sigma\sqrt{2\pi e})$. A raíz de este resultado fundamental se puede interpretar que los errores que dan lugar a una desviación sobre la distribución gaussiana han de tener información física esencial para comprender el fenómeno estudiado.

6.2 Desviaciones sobre la gaussiana

Cuando se realiza un análisis espectral para encontrar las periodicidades de una serie temporal la herramienta más extendida es el periodograma (ver Cap. 2.2). Para la detección de frecuencias se utiliza el test de Fisher (Fisher, 1929) que se obtiene a través de la teoría de distribuciones. Este test se basa en que las variables sean aleatorias independientes e idénticamente distribuidas (i.i.d.) y gaussianas de modo que si X_j es un proceso gaussiano su transformada de Fourier

$$S_n(\phi) = \sum_{j=1}^n X_j \exp(ij\phi)$$

6. CONSISTENCIA DEL ANÁLISIS

es un proceso gaussiano complejo, y partiendo de las propiedades de la distribución gaussiana se puede determinar la significancia de un determinado pico en el espectro de Fourier. En cambio, si el proceso X_j no tiene una distribución gaussiana la distribución de S_n no tendrá una forma cerrada y no se podrá determinar la significancia de los picos.

Entendemos las curvas de luz estelares como series temporales compuestas por 2 procesos, uno que representa a la señal y otro que representa al ruido. El primero es una función de tipo armónico y el segundo es no armónico. La distribución de una función armónica sencilla no es de tipo gaussiano pero en la práctica, la distribución de los datos procedentes de las observaciones están dominados por una componente con este tipo de distribución que es la que se interpreta como ruido. Esta componente aparece como residuo al hacer un proceso de *prewhitening* de las componentes armónicas detectadas en la serie temporal.

A menudo al analizar las observaciones por satélite de estrellas pulsantes surgen distribuciones que no son de tipo gaussiano (véase p.ej. [Degroote et al. 2009](#)) pero cuando las desviaciones son pequeñas se supone que los métodos clásicos pueden seguir usándose. No obstante, la estimación de los parámetros (las amplitudes en la ref. citada) puede verse sesgada debido a una asimetría estadística. Para determinar los parámetros bajo la hipótesis del test de Fisher en situaciones en las que la distribución de los errores es no gaussiana se introducen correcciones que tienen en cuenta efectos de correlación que describen estadísticamente la desviación sobre la gaussiana. Sin embargo, no se proporciona un modelo físico de estas correlaciones y la no gaussianidad se desprecia como una simple perturbación.

Entre las posibles causas de no gaussianidad en los residuos se halla la presencia de efectos no lineales en la curva de luz. Se puede demostrar por el Teor. Central del Límite que, dada una realización de un proceso de tipo gaussiano $X(t)$ hasta un tiempo t' determinado, la media condicional (valor más probable de una predicción) del valor $X(t' + \tau)$ con $\tau > 0$ es una función lineal de $\{X(t); t \leq t'\}$. Este resultado justifica el análisis lineal de este tipo de procesos, de modo que bajo la hipótesis de gaussianidad se suele restringir la atención a funciones lineales. Pero este resultado nos indica además, que la falta de linealidad en el proceso puede contraer que la distribución sea no-gaussiana.

Existen varias explicaciones para la aparición de no linealidades en curvas de luz como son: el acoplamiento a través de resonancias entre modos que interactúan ([Dziembowski, 1982](#)), las oscilaciones en regímenes altamente no-adiabáticos ([Saio et al., 1998](#)), o una distorsión en la curva de luz causada por una respuesta no lineal del flujo estelar ([Garrido and Rodríguez, 1996](#)).

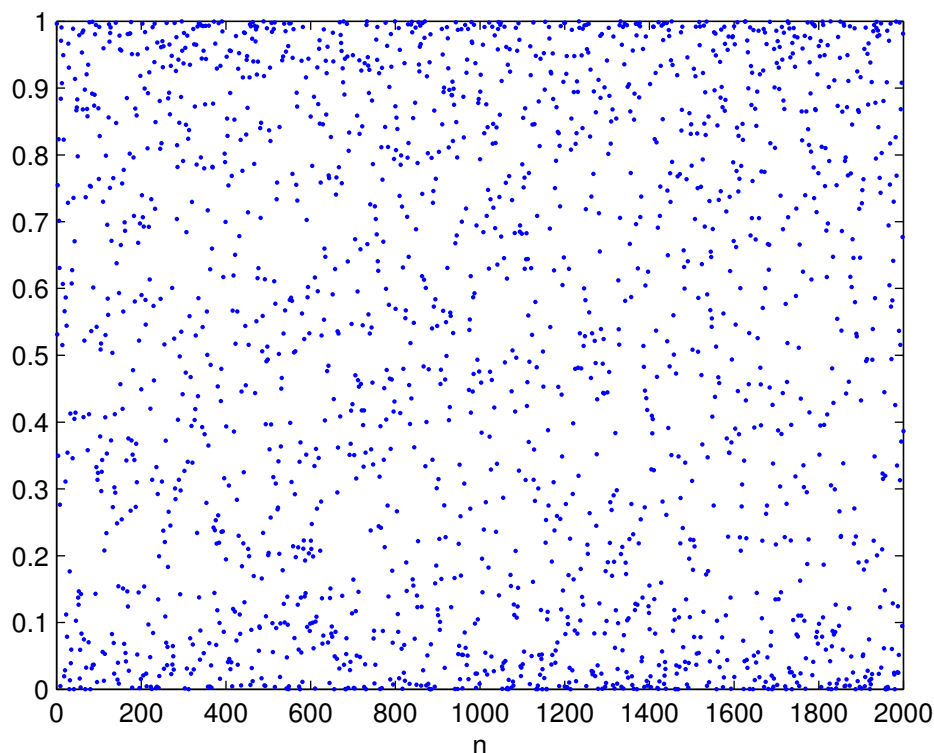


Figura 6.2: Serie de datos aleatorios generados según una distribución de tipo $\mathcal{B}(0.5, 0.5)$ en el intervalo $[0, 1]$. Nótese como los datos se apilan en los extremos del intervalo.

Otra causa de no gaussianidad es la falta de estacionaridad de la curva de luz, es decir la variación temporal de los momentos de la distribución (Priestley, 1988). Esto puede manifestarse como una variación de amplitudes y fases con el tiempo que podría estar en relación también con los propios efectos de la no linealidad de la serie.

En cualquier caso, cualquier teorización debe ajustarse a los hechos y no a la inversa, de modo que si no poseemos información adicional no podemos descartar cualquier otro origen para la aparición de efectos no lineales en las curvas de luz.

Finalmente, existe una causa de no gaussianidad no considerada en la literatura pero que puede probarse matemáticamente, esta es la no analiticidad de la función. En efecto, puede demostrarse (Ventkus, 1982) que una función aleatoria gaussiana es infinitamente diferenciable y analítica. Por tanto, una función no analítica ha de ser no gaussiana. Esto resulta evidente cuando se comparan la Fig. 6.1 y la Fig. 6.2.

Se puede apreciar que, a diferencia de lo que ocurre con una secuencia aleatoria con distribución gaussiana, los datos generados mediante una distribución no gaussiana como la $\mathcal{B}(0.5, 0.5)$, es decir, una distribución Beta con parámetros $\alpha = \beta = 0.5$, presentan saltos que son propios de una función con problemas de diferenciableidad. Este es un ejemplo claro de una distribución no gaussiana que es no analítica.

Este aspecto, el de la no analiticidad de las series, es el que nos concierne en el

siguiente apartado y en los próximos capítulos pues, además de ser una posible causa de la no gaussianidad de las distribuciones de residuos, complica el análisis armónico que se lleva a cabo para la detección de frecuencias. De hecho, como veremos, un criterio para la consistencia del análisis de frecuencias realizado mediante técnicas de Fourier es que la función que subyace a las series temporales sea analítica. Si esta condición no se cumple, no está garantizada la consistencia del análisis.

6.3 Condiciones para el análisis armónico

Para poder aplicar las técnicas de Fourier a una función dada, deben satisfacerse las condiciones para que se cumpla el Teor. de Parseval (Kaplan, 1992). Este teorema dice que la integral del módulo al cuadrado de una función es igual a la integral del módulo al cuadrado de su transformada de Fourier. Para que esto se cumpla la función debe ser de cuadrado integral, es decir,

$$\int_{-\infty}^{+\infty} |F(x)|^2 dt < \infty \quad (6.2)$$

La expansión de Fourier de una función converge puntualmente c.t.p. cuando ésta es de cuadrado integrable (Carleson, 1966). A diferencia de las series de Fourier, que son una suma infinita, la Transformada de Fourier Discreta (TFD) y su inversa, que es la análoga a las series de Fourier, siempre convergen a un valor si la secuencia es finita, dado que la suma es finita. Pero cuando la serie de Fourier de la función es divergente o no converge al valor de la función en cada punto, la TFD deja de proporcionar una descripción física de la función basada en sus componentes de frecuencia. Es decir, una frecuencia es sólo una variable física bien definida cuando la serie de Fourier converge.

Por lo tanto, antes de que se pueda aplicar un análisis de Fourier a una serie temporal se debe demostrar la convergencia de la expansión en series de Fourier de la función.

Por otro lado, la analiticidad de una función es condición suficiente para que se cumpla la integrabilidad del cuadrado (van Dijk, 2009). Se dice que una función es analítica si es infinitamente diferenciable de modo que su expansión en serie de Taylor es convergente. Si esta condición no se cumple la convergencia de la serie de Fourier no está garantizada. Esta es la razón por la que es tan importante evaluar la analiticidad de la función antes de llevar a cabo el análisis armónico. De aquí en adelante nos referiremos a analiticidad como la propiedad de que una función pueda ser desarrollable en una serie de Fourier. Para estudiar la analiticidad de una función introduciremos en el Cap. 7 una aproximación para hallar la diferenciabilidad de series discretas: la conectividad.

Capítulo 7

El método de las conectividades

*“It has long been an axiom of mine that the little things
are infinitely the most important”*

The Adventures of Sherlock Holmes - A. C. Doyle

En este capítulo de la memoria introducimos una metodología que, evaluando la analiticidad de la función que subyace a las observaciones, permite verificar que se cumplen las condiciones necesarias para el análisis armónico. La analiticidad de las curvas de luz puede estar, además, en conexión con la no gaussianidad encontrada en la mayoría de ellas como hemos visto en el Cap. 6.2.

7.1 Conectividad

7.1.1 Concepto

En Astronomía es frecuente que los datos observacionales consistan en una secuencia discreta de muestreos de una función subyacente, que generalmente se considera continua¹. Esta función es la representación de un observable físico dado. Entendemos por observable la cantidad física que satisface todas las propiedades derivadas de la definición de medida de Lebesgue (Jech, 1997). Es importante destacar aquí que un observable de una propiedad física dada no tiene por qué coincidir necesariamente con la función que lo representa.

La secuencia discreta de variaciones fotométricas de la luminosidad de las estrellas pulsantes es lo que se conoce como curva de luz. Podemos estudiar la variabilidad estelar a través de la curva de luz o a través de la secuencia de medidas de la velocidad radial (RV).

¹Cuando hablamos aquí de función continua nos estamos refiriendo a una secuencia de datos espaciados infinitesimalmente, y por discreta nos referimos a una secuencia de datos con un espaciado finito.

7. EL MÉTODO DE LAS CONECTIVIDADES

Al considerar la función continua que representa al observable la primera cuestión que surge es: ¿podemos hallar una secuencia discreta de valores que determine completa y únicamente las propiedades de la función entera sin necesidad de conocer los infinitos valores de esta? Este es el clásico problema del muestreo y para un función de banda limitada el Teor. de Nyquist-Shannon ([Shannon, 1949](#)) establece que la función se puede reconstruir perfectamente a través de un conjunto finito de datos de la función con una separación equidistante si la función es de banda limitada y la anchura de esta es menor que la mitad de la frecuencia de muestreo.

Cuando elegimos las condiciones de forma que la respuesta a la pregunta de antes es afirmativa, como es el caso de los datos procedentes de misiones como *CoRoT* o *Kepler*, deberíamos poder estudiar todas las propiedades de la función subyacente incluso las que conciernen a los límites en un punto de la función. Sobra decir que excluimos de esta argumentación los eventos singulares que no son propios de la señal sino de artefactos.

Dicho todo esto, estamos interesados ahora en una propiedad que se relaciona con la forma en que los puntos de una secuencia discreta pueden ser extrapolados para reproducir con exactitud la función continua subyacente (en adelante omitiremos el adj.subyacente). Esta propiedad, que está intrínsecamente relacionada con la analiticidad de la función, es lo que llamamos la conectividad de los datos.

Normalmente para estudiar la continuidad de una función en un punto hallamos los límites laterales y comprobamos que convergen al mismo valor y que este es el valor de la función en ese punto. De una forma similar, dado un punto de una secuencia discreta podemos calcular una extrapolación adelante usando los datos que le preceden, y a la vez una extrapolación atrás a partir de los datos posteriores, y comprobar si ambas extrapolaciones convergen hacia el mismo valor y si coincide con el valor del punto de la secuencia seleccionado.

Podemos generalizar diciendo que la secuencia discreta generada a partir de la función continua es una secuencia estocástica con una parte determinista que puede tener una estructura de tipo oscilatorio, que es precisamente la señal en la que estamos interesados. De hecho, es el Teor. de descomposición de Wold ([Wold, 1938](#)) el que nos asegura que esto sea así al establecer que cada proceso estacionario puede ser representado siempre mediante un proceso estocástico que posee una componente puramente determinista como se comentó en la primera parte de la memoria.

Ahora bien, la componente aleatoria que contiene el proceso estocástico podría complicar el análisis de la conectividad de los datos debido a que las señales de tipo aleatorio no son continuas en un sentido estricto, pero la generalización anterior nos permite hacer uso de un importante teorema matemático para demostrar la consistencia del método que proponemos para estudiar la analiticidad de la función subyacente.

El Teor. de Continuidad de Kolmogorov (Revuz and Yor, 1999), que se aplica a series estocásticas, permite extender propiedades a toda la serie a partir de un conjunto finito de datos. Dicho de otro modo, según este teorema, propiedades como la diferenciabilidad, cuya determinación requiere conocer datos infinitamente cercanos y no pueden ser determinadas mediante las definiciones clásicas en procesos estocásticos, pueden extenderse a partir de un número finito de muestreos de la serie estocástica. La extensión de las propiedades es posible siempre que se satisfagan ciertas condiciones sobre los momentos estadísticos de los incrementos de los datos que garantizan la separabilidad.

Básicamente estas condiciones se cumplen cuando los momentos de los incrementos varían como mucho como una ley de potencias. Ya que esto es lo esperado para el tipo de datos que usamos en astrosismología, estas propiedades pueden estudiarse en las series temporales que analizamos a pesar de la componente aleatoria. No se espera, por tanto, que la componente aleatoria de los datos complique la caracterización de la función continua a través de las conectividades.

7.1.2 Definición

La argumentación anterior nos asegura la consistencia del método descrito de forma cualitativa más arriba. Ahora vamos a una definición más formal de la propiedad que hemos llamado conectividad \mathcal{C}_n de un punto X_n de la secuencia discreta muestreada de la función $F(x)$.

$$\mathcal{C}_n = \epsilon_n^f - \epsilon_n^b, \quad (7.1)$$

donde X_n^f y X_n^b son las predicciones por delante y detrás del punto X_n respectivamente, y ϵ_n^f , ϵ_n^b , son los errores de dichas predicciones.

$$\begin{aligned} \epsilon_n^f &= x_n^f - x_n \\ \epsilon_n^b &= x_n^b - x_n. \end{aligned} \quad (7.2)$$

A partir de esta definición se puede expresar una aproximación a la condición de derivabilidad en un punto para $F(x_n)$:

$$\frac{x_n^b - x_{n+1}}{\Delta t} = \frac{x_n^f - x_{n-1}}{\Delta t} \quad (7.3)$$

donde Δt es el intervalo de muestreo de la secuencia. Es fácil ver a partir de estas ecuaciones que la conectividad es una propiedad intrínsecamente relacionada con el

7. EL MÉTODO DE LAS CONECTIVIDADES

concepto de derivabilidad. La aproximación numérica a la derivada en un punto \mathcal{D}_n en x_n puede expresarse como:

$$\mathcal{D}_n = \frac{\mathcal{C}_n + x_{n+1} - x_{n-1}}{2\Delta t} \quad (7.4)$$

que se reduce a la típica derivada en un punto (Burden and Faires, 2011) para datos discretos cuando $\mathcal{C}_n = 0$. Cuando las conectividades no son cero sino valores distribuidos aleatoriamente de manera independiente, la derivada todavía puede considerarse bien definida. En ese caso las conectividades pueden considerarse simplemente las desviaciones de la derivada en ese punto. De lo contrario, la condición de derivabilidad no se cumpliría.

7.1.3 El coeficiente de no-diferenciabilidad de una función

En su artículo de 1923 titulado “Differential Space”, Norbert Wiener (1923) aborda el estudio de procesos estocásticos gaussianos como el movimiento browniano a través del funcional que describe la trayectoria en un espacio que se considera de dimensión infinita por ser esta muy grande. Sin embargo, el espacio que él llama espacio diferencial no es el espacio de posiciones sino el de las cantidades

$$\begin{aligned} x_1 &= f\left(\frac{1}{n}\right) - f(0) \\ x_2 &= f\left(\frac{2}{n}\right) - f\left(\frac{1}{n}\right) \\ \dots x_k &= f\left(\frac{k}{n}\right) - f\left(\frac{k-1}{n}\right) \\ x_n &= f(1) - f\left(\frac{n-1}{n}\right). \end{aligned}$$

donde los $f\left(\frac{1}{n}\right), \dots, f\left(\frac{k}{n}\right), \dots, f(1)$ son las dimensiones del funcional $f(t)$ del proceso que se estudia. Así pues, Wiener obtiene el siguiente resultado fundamental:

La probabilidad de que

$$\lim_{n \rightarrow \infty} \sum_1^n \left(f\left(\frac{k}{n}\right) - f\left(\frac{k-1}{n}\right) \right)^2 = 0 \quad (7.5)$$

para una función continua de variación total limitada T sobre el intervalo $[0, 1]$, y en particular, cuando $f(t)$ posee una derivada acotada en ese intervalo cerrado.

De este modo, el límite de arriba puede considerarse como una especie de coeficiente de no-diferenciabilidad de f .

Ahora bien, volviendo al caso que nos ocupa, el funcional $f(t)$ se puede considerar en nuestro caso un vector $\|X\|$ de valores en tiempos discretos de las observaciones,

donde $X(t)$ es nuestra serie temporal. En [Schwarzenberg-Czerny \(2003\)](#), por ejemplo, Schwarzenberg-Czerny estudia la estadística de las series temporales en astronomía a partir de la ortogonalidad de los componentes de ese vector n -dimensional $\|X\|$. Siguiendo la analogía con el funcional de Wiener, podríamos considerar que el resultado anterior se puede aplicar a nuestras series temporales también para estudiar la diferenciabilidad de estas.

Así pues, nuestro test incluye además del cálculo de las conectividades el estudio de la tendencia asintótica de la suma de los cuadrados de las \mathcal{C}_n con el número de puntos n usados para calcular las conectividades.

Cabe mencionar, que el artículo de Wiener, es un resultado que se desconocía a priori a la hora de realizar este trabajo de tesis, pero se hace una mención especial en éste apartado dado el paralelismo de los resultados a los que se ha llegado de manera independiente.

Ahora para calcular las conectividades, las extrapolaciones requieren de un modelado de los datos de forma paramétrica. Para ello hacemos uso de 2 aproximaciones que veremos en el próximo apartado: los splines y métodos autoregresivos de media móvil (ARMA).

7.2 Métodos

7.2.1 Splines

Para hallar las predicciones hacia adelante y hacia atrás tenemos que usar un modelo que pueda aproximar arbitrariamente bien las funciones continuas. El Teor. de Stone-Weierstrass ([Royden, 1988](#)) establece que una función uniformemente continua en un intervalo cerrado $[a,b]$ puede ser aproximada arbitrariamente bien, es decir, con toda la precisión que se desee, por un polinomio de grado n , donde n es un número natural. Entonces, la función continua entera puede expresarse como una función polinómica a trozos. No obstante, las funciones polinómicas a trozos tienen un problema, el fenómeno de Runge, que se caracteriza por mostrar unas oscilaciones en los extremos de una función, esto es similar a lo que ocurre con el fenómeno de Gibbs en la aproximación de Fourier.

Otro punto crucial aquí es que los polinomios a trozos son funciones continuas que no necesariamente tienen que ser analíticas ya que pueden tener problemas con las derivadas de orden alto. Entonces, en lugar de elegir los polinomios para modelar la función, elegimos **splines**, es decir, una función definida a trozos de polinomios suaves. Esta condición adicional que es la suavidad de la función, garantiza la analiticidad de la extrapolación y por tanto la convergencia de su desarrollo en serie de Fourier. Al

7. EL MÉTODO DE LAS CONECTIVIDADES

mismo tiempo, se resuelve el problema del fenómeno de Runge ya que este desaparece al usar funciones suaves.

Definido bajo estas condiciones el problema se presenta de forma muy general de modo que todas las consecuencias derivadas no sólo se aplican funciones polinómicas. Si podemos ajustar arbitrariamente bien la función continua con un modelo paramétrico de splines la condición para el análisis armónico que es la condición de analiticidad, se cumple. De lo contrario, la convergencia de las series de Fourier no está garantizada para la función continua.

La aproximación con splines nos da una información esencial sobre propiedades local de la función continua que genera la secuencia discreta observada como son la diferenciabilidad. Pero podemos obtener una percepción más profunda sobre el origen físico de dichas propiedades si comparamos los resultados con los que se obtienen al hacer uso de otra aproximación que no se base en funciones analíticas, es decir, que sea capaz de ajustar la función continua también en el caso de que esta no sea analítica. Los modelos autoregresivos, aunque son de difícil interpretación, nos proporcionan esa base ya que el ajuste que proporcionan siempre puede reducir los datos a una serie de valores aleatorios estrictamente independientes, de modo que modelizan todas las propiedades presentes en los datos.

7.2.2 ARMA

La segunda aproximación que usamos para la predicción hacia adelante y hacia atrás de cada punto para calcular las conectividades son modelos ARMA (Box and Jenkins, 1976) (estos fueron definidos anteriormente en el Cap. II). Al igual que con el modelado de splines, la precisión de las extrapolaciones que se llevan a cabo mediante modelos ARMA depende del número de datos del segmento modelado. Pero además, las extrapolaciones ARMA dependen de 2 parámetros: los órdenes p y q , que son el número de coeficientes autoregresivos y el número de coeficientes de media móvil, respectivamente.

Por lo general, un modelo ARMA puede reproducir una función seno con sólo 3 términos y conforme el número de componentes de frecuencia aumenta el número de términos necesarios del modelo ARMA para reproducir la señal aumenta de manera correspondiente.

El algoritmo usado para calcular los modelos ARMA se basa en una minimización de los residuos y por tanto, es insensible a la flecha del tiempo (Scargle, 1990). Sin embargo, extrapolamos sólo un punto cada vez y los segmentos de datos modelados son suficientemente cortos como para poder estimar las derivadas sin que la dirección del tiempo introduzca un sesgo en los resultados. Por otro lado, la invariancia a la flecha del tiempo nos permite calcular las extrapolaciones hacia atrás simplemente

invirtiendo la matrix de la extrapolación hacia adelante para el segmento de datos posteriores al punto en cuestión.

Para hallar las extrapolaciones ARMA usamos un algoritmo iterativo que involucra los siguientes pasos:

1. Identificación del orden inicial del modelo - afecta principalmente a la velocidad de convergencia.
2. Cálculo de los parámetros usando un algoritmo de optimización dados los órdenes para las contribuciones AR y MA del modelos.
3. Evaluación de la validez del modelo basada en la bondad del ajuste en las extrapolaciones hacia atrás y adelante. El algoritmo empieza de nuevo con los nuevos valores de los órdenes incrementados hasta que se alcanza un mínimo en los residuos.

Cabe mencionar aquí, finalmente, que a pesar de tener una difícil interpretación física los modelos ARMA tienen una serie de ventajas sobre los modelos que parten de una descomposición en funciones de forma - cerrada. Al contrario de técnicas como las de Fourier, los métodos autoregresivos no usan ninguna forma de expresión analítica. Así pues, la descomposición de Fourier es útil para la descripción de procesos periódicos con una variación no-periódica pequeña ya que la expansión en series de Fourier de una función periódica es finita. Pero cuando estamos modelando todos los aspectos de una función y algunos de esos aspectos son rasgos no-periódicos cuya expansión en series de Fourier puede ser infinita, puede ser más apropiado el uso de métodos autoregresivos.

En el próximo capítulo nos dedicaremos a aplicar los métodos descritos aquí a casos de estudio en astrosismología.

7. EL MÉTODO DE LAS CONECTIVIDADES

Analiticidad de las series observadas

“I turn away with fear and horror from the lamentable plague of continuous functions which do not have derivatives...”

Hermite, letter to Stieltjes, 20 de mayo de 1893

Probamos a continuación la metodología introducida en el capítulo anterior usando datos obtenidos mediante distintos satélites (CoRoT, *Kepler*, y SoHO) para estrellas pulsantes de diferentes características (Pascual-Granado et al., 2014). Concretamente estudiamos la analiticidad de las curvas de luz de las siguientes estrellas:

Nombre	Tipo de pulsante	Tipo de medida	Instrumento	Referencia
HD 174936	δ Sct	fotometría	CoRoT	García Hernández et al. (2009)
KIC 006187665	híbrida δ Sct/ γ Dor	fotometría	<i>Kepler</i>	Uytterhoeven et al. (2011)
El Sol	solar-like	vel. radiales	SoHO/GOLF	García et al. (2005)
HD 181072	Constante	fotometría	CoRoT	Degroote et al. (2009)

Tabla 8.1: Resumen de las estrellas analizadas.

Para la estrella HD 174936 hemos realizado además, el test de analiticidad sobre un modelo analítico que construimos basándonos en el contenido en frecuencias publicado en García Hernández et al. (2009). Comparando los resultados con el modelo analítico hallado mediante técnicas clásicas de Fourier podemos comprobar si propiedades de la curva de luz original como la analiticidad son reproducibles en el modelo así construido, evaluando de este modo la validez del desarrollo en series de Fourier de las curvas de luz.

8.1 HD 174936

Esta estrella representa un caso paradigmático de los retos que surgen al tratar de encontrar una interpretación física de las curvas de luz estelares mediante un análisis de Fourier. En [García Hernández et al. \(2009\)](#), con un límite bastante conservador en señal-ruido, encontraron que la curva de luz de esta estrella presenta 422 frecuencias significativas. Sin embargo, ninguno de los modelos estudiados predice el intervalo completo de frecuencias hallado en las observaciones mediante técnicas de Fourier.

Entre las distintas hipótesis propuestas para explicar esta discordancia entre teoría y observaciones, Kallinger ([Kallinger and Matthews, 2010](#)) propuso que la mayoría de los picos en el espectro de Fourier son causados por un ruido coloreado de fondo debido a la granulación, y determinaron que menos de 100 frecuencias son debidas a pulsaciones después de eliminar una función de ley de potencias al espectro.

Este procedimiento parece reproducir bien la densidad de frecuencias que se esperan teóricamente pero no deja de ser una hipótesis que no tiene una relación contrastada con el origen físico de las frecuencias y, de hecho, no existe un modelo de la granulación para este tipo de estrellas que justifique usar dicho procedimiento.

Dadas las peculiaridades del espectro de esta estrella, y puesto que es una estrella que se ha estudiado bien en la literatura, se ha seleccionado como caso de estudio principal para nuestras pruebas. No obstante, es importante destacar aquí que este es sólo uno de los múltiples casos en que los análisis de objetivos observados por los satélites CoRoT y *Kepler* han mostrado muchas más frecuencias de las esperadas.

8.1.1 Modelo analítico para HD 174936

Este modelo se construye usando las amplitudes, frecuencias y fases de las 422 frecuencias halladas mediante técnicas clásicas de Fourier y ajustadas a una expansión en series de Fourier de la forma:

$$x^a(t_i) = \sum_{i=1}^n A_i \cos(2\pi \nu_j t_i + \phi_j) + N^b \quad (8.1)$$

donde A_i , ν_j , y ϕ_j representan respectivamente las amplitudes observadas, las frecuencias, y las fases. El término N^b representa el ruido blanco definido como una distribución normalizada escalada a los datos teniendo en cuenta la proporción señal-ruido de [García Hernández et al. \(2009\)](#)

$$N^w = \langle \sigma^N \rangle \quad (8.2)$$

En la Fig. [8.1](#) comparamos la curva de luz original de CoRoT con una realización del modelo analítico que acabamos de describir.

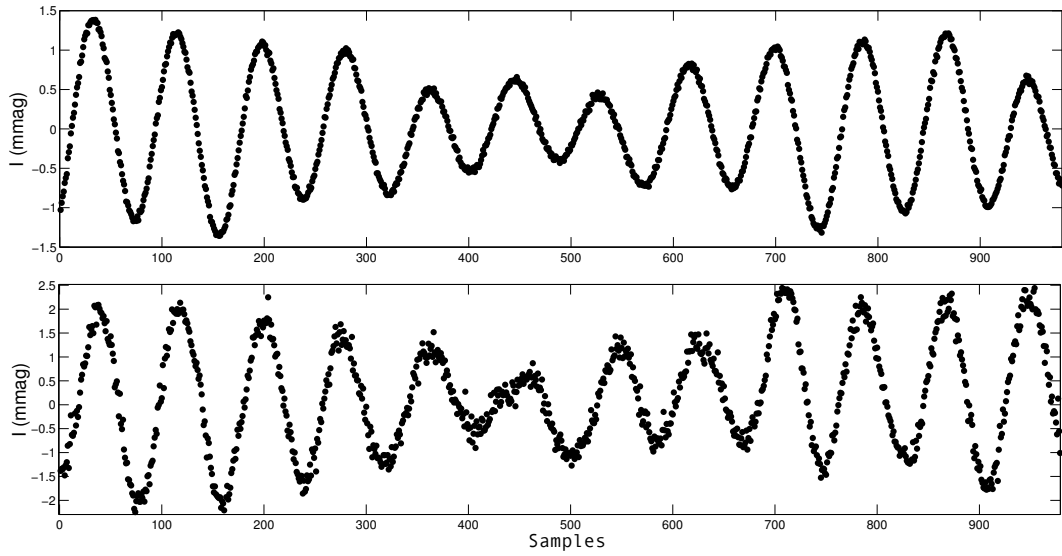


Figura 8.1: Segmento de 1000 puntos de la curva de luz de la estrella HD 174936. El panel superior muestra una realización del modelo analítico que ajusta la curva de luz de esta estrella. El panel inferior muestra los datos originales tomados por CoRoT.

Además de este modelo hemos definido también un segundo modelo analítico en el que añadimos al término de ruido blanco una componente de ruido coloreado. Así pues el segundo modelo analítico queda expresado como

$$x^a(t_i) = \sum_{i=1}^n A_i \cos(2\pi \nu_j t_i + \phi_j) + N^b + N^{\text{nb}} \quad (8.3)$$

y la componente de ruido coloreado se calcula como

$$N^{\text{nb}} = \mathcal{F}^{-1}[\tilde{N}^{\text{nb}}] \quad (8.4)$$

donde \tilde{N}^{nb} se define en el dominio transformado como

$$\tilde{N}^{\text{nb}} = \left[\left| \sum_{k=1}^m \frac{a_k}{1 + \left(\frac{\nu}{b_k}\right)^{c_k}} \right| \right]^{1/2} e^{i\theta} \quad (8.5)$$

siendo a_k y b_k parámetros libres que se ajustan. Los valores de $c_k = 4$ y $m = 2$ se adoptaron siguiendo a [Kallinger and Matthews \(2010\)](#). Este ruido se calculó haciendo las fases θ igual a una distribución aleatoria uniforme en el intervalo $[-\pi, \pi]$. La curva de luz así construida se supone originada por los modos normales de oscilación de la estrella pulsante mas el ruido de granulación modelado como turbulencia convectiva ([Harvey, 1985](#)).

8.1.2 Análisis de las conectividades

Como se explicó en el Cap. 7 el primer paso para el modelado ARMA consiste en la identificación de los órdenes del modelo. Tras iterar el algoritmo hasta obtener los errores de predicción mínimos obtuvimos que $p = 20$ y $q = 1$ son los órdenes óptimos para modelar la curva de luz de HD 174936. Entonces calculamos las conectividades mediante el modelo ARMA(20,1) y mediante polinomios splines. Los datos de CoRoT tienen huecos causados por el paso del satélite por la Anomalía Sur Atlántica (SAA) que ocurre 2 veces al día (Samadi et al., 2007). Para evitar hacer correcciones seleccionamos un segmento de la curva de luz de 1000 puntos (0.37 días) que carece de huecos y *outliers* y lo normalizamos estadísticamente. A continuación hallamos las conectividades para cada punto modelando 80 datos cada vez: 40 puntos para las extrapolaciones adelante y 40 puntos para las extrapolaciones hacia atrás.

Como puede verse en las Figs. 8.2 y 8.3 las conectividades spline tienen una dispersión mucho mayor que las conectividades ARMA, con un efecto más dramático en el caso de los datos de CoRoT. Por otro lado, a pesar de los valores grandes de las conectividades spline para el modelo analítico, no se ha podido hallar una correlación con los datos originales de modo que se puede suponer que los valores encontrados se deben al error de predicción de la aproximación spline en este caso, que disminuye cuando el número de puntos usados durante el modelado aumenta. El aspecto más destacable sobre las conectividades splines resulta en el caso de los datos originales de CoRoT. En ese caso éstas presentan una clara correlación con los datos originales (cf. Fig. 8.2). El mismo efecto puede ser observado también en las conectividades ARMA pero la escala es 5 veces menor. Este resultado sugiere que el comportamiento de estos datos no puede ser modelado por polinomios spline.

Los resultados del análisis de conectividades para el segundo modelo analítico definido en el apartado anterior se muestran en la Fig. 8.4. En este caso hemos superpuesto las 3 series sobre un mismo gráfico para que se puedan apreciar mejor las diferentes escalas. De este modo es posible notar que añadiendo una componente de ruido coloreado al modelo los resultados obtenidos para las conectividades no se ajustan mejor a los que se obtienen con los datos originales. De hecho, las conectividades se reducen respecto al primer modelo estudiado, indicando que el segundo modelo aún se ajusta peor al análisis de conectividades de los datos originales. Esto es comprensible ya que, por lo general, las fluctuaciones que introduce un ruido coloreado son más suaves que las del ruido blanco y por tanto deben afectar menos a la diferenciabilidad.

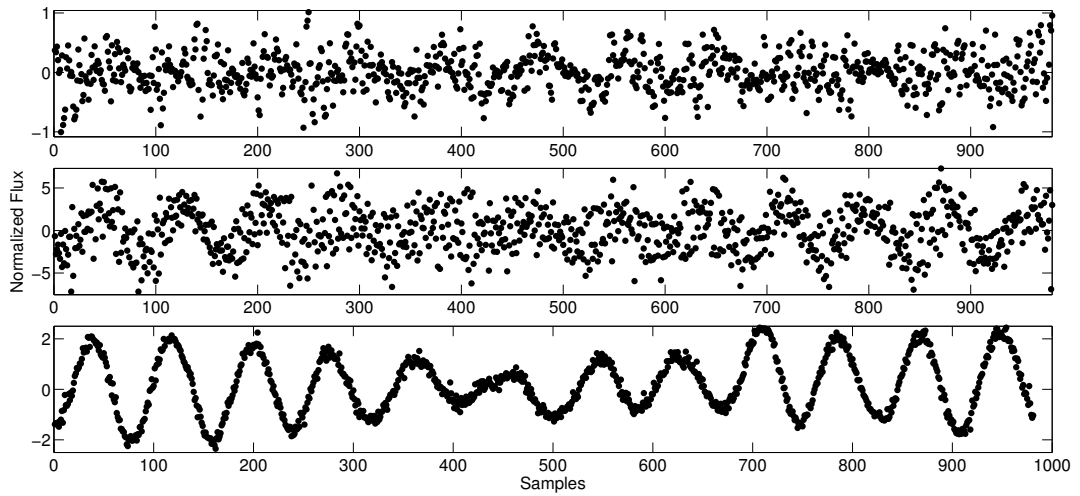


Figura 8.2: Conectividades de la curva de luz de la estrella HD 174936. En el panel superior se muestran las conectividades calculadas con la aproximación ARMA, en el panel intermedio se muestran las conectividades spline, y en el panel inferior los datos originales de CoRoT para su comparación. Nótese las diferentes escalas.

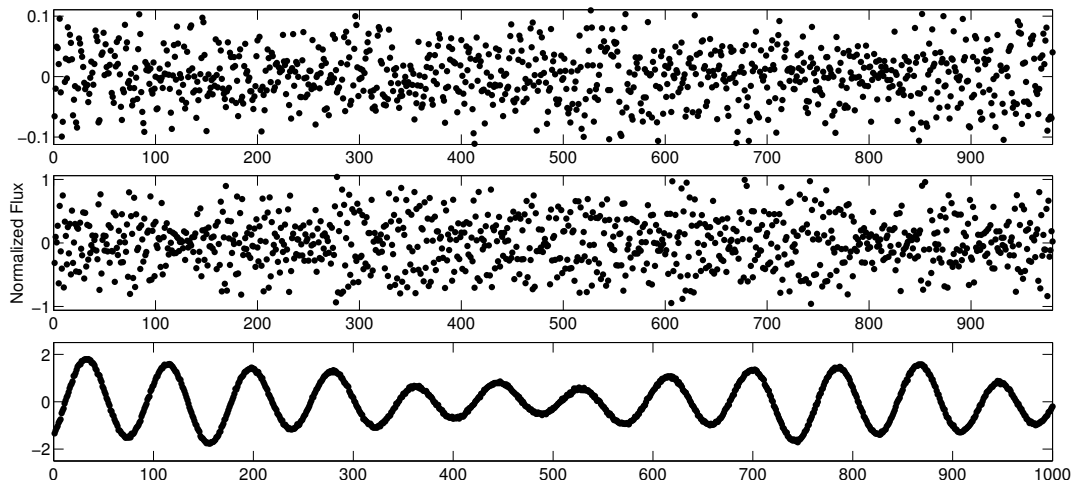


Figura 8.3: Conectividades para el modelo analítico-1 de la curva de luz de HD 174936. En el panel superior se muestran las conectividades calculadas con la aproximación ARMA, en el panel intermedio se muestran las conectividades spline, y en el panel inferior una realización del modelo analítico para su comparación. Nótese las diferentes escalas.

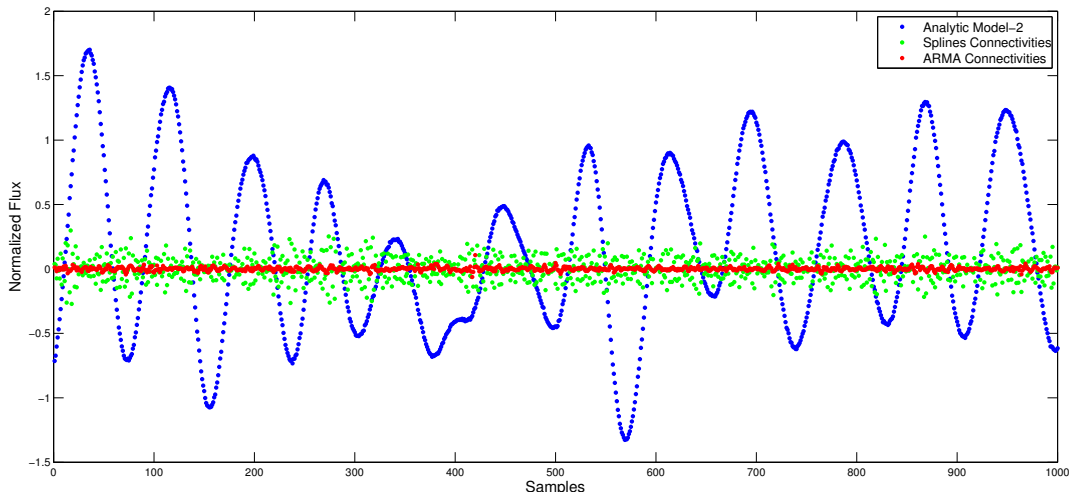


Figura 8.4: Conectividades para el modelo analítico-2 de la curva de luz de HD 174936. En rojo se muestran las conectividades calculadas con la aproximación ARMA, en verde se muestran las conectividades spline, y en azul los puntos correspondientes a una realización de este modelo.

8.1.3 SSE

Estudiamos ahora como varía la suma del cuadrado de las conectividades - aquí llamada SSE (del inglés “*Sum of the Square Errors*”) por la similitud de éstas con los residuos - con el número de puntos suministrados al modelo que se ajusta. Es decir, evaluamos la tendencia asintótica de la SSE con n .

La precisión de las predicciones depende del número de puntos usados para hallar los modelos, de forma que para un modelo que se puede ajustar arbitrariamente bien, como es el caso de los splines cuando se ajustan funciones analíticas (ver Cap.7.2.1), se espera que esta suma tienda a cero o a un valor finito cercano a cero. Por tanto, este coeficiente que llamamos SSE está intrínsecamente relacionado con el coeficiente de no-diferenciabilidad de una función descrito por Wiener (1923).

En la Fig. 8.5 estudiamos como este coeficiente cambia con el número de puntos (N_p) usados para el modelado en cada caso: conectividades spline o ARMA, y modelo numérico o datos de CoRoT. La curva empieza en $N_p = 30$ ya que con menos datos no es posible calcular el modelo ARMA(20,1) usado para este estudio.

Lo primero que puede notarse en el gráfico correspondiente a los datos de CoRoT (panel superior) es la tendencia asintótica hacia cero del SSE calculado mediante la aproximación ARMA. Esto significa que suministrando suficientes datos al modelo ARMA(20,1) este ajusta adecuadamente la serie de datos. En el caso de la aproximación splines la tendencia asintótica es hacia un valor finito significativamente distinto de cero además de mostrar una serie de oscilaciones que parecen indicar que no se están ajustando todas las variaciones periódicas de la señal. Así pues, para los

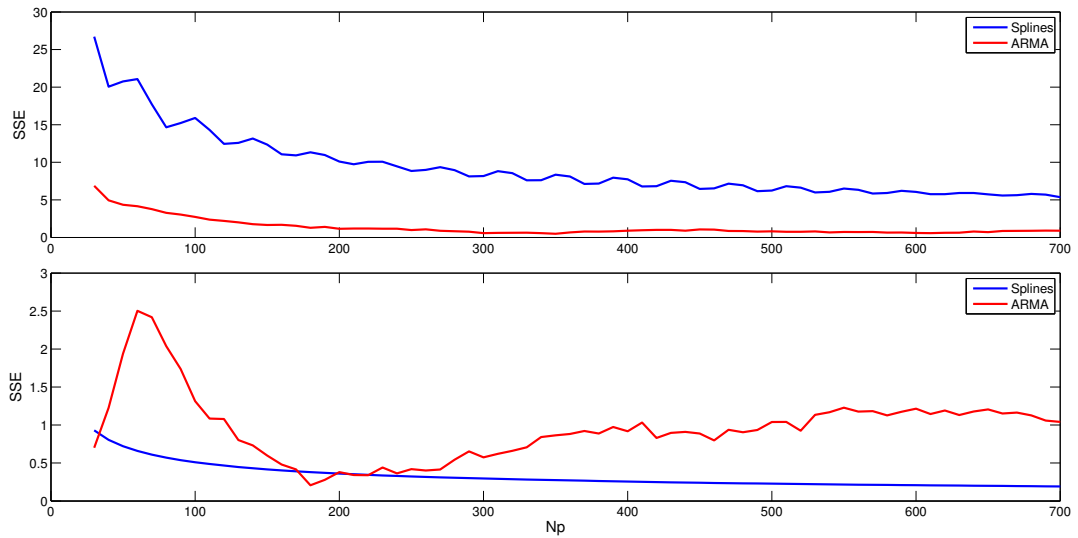


Figura 8.5: Tendencia asintótica del coeficiente SSE con el número de puntos N_p a partir de las conectividades ARMA (en rojo), y las conectividades spline (en azul) para los datos de CoRoT (panel superior) y para el modelo analítico-1 (panel inferior).

datos de CoRoT la aproximación ARMA que ajusta funciones no analíticas funciona mucho mejor que los splines, que sólo ajustan funciones analíticas. Esto indica que la función que subyace a la serie de CoRoT no es una función analítica.

En el panel inferior podemos apreciar que los resultados del SSE para el modelo analítico son muy distintos a los de la serie de CoRoT. La aproximación de splines tiende a un valor cercano mucho más cercano a cero para el modelo analítico porque puede ajustar arbitrariamente bien una función que no tenga problemas de diferenciabilidad, y este modelo se ha construido con funciones de este tipo. Por otro lado, la SSE que el modelo ARMA(20,1) produce evoluciona de forma irregular y no tiende en absoluto hacia cero. Esto es de esperar teniendo en cuenta el contenido en frecuencias de este modelo. En efecto, si un modelo ARMA(2,0) representa función armónica sencilla como un coseno, entonces un modelo ARMA(20,1), que tiene 21 parámetros sólo puede ajustar pobremente una función con 422 frecuencias presentes, como es el caso de nuestro modelo analítico. En definitiva, el modelo analítico construido es una función con un contenido en frecuencias que no puede ajustar el modelo ARMA que mejor ajusta los datos originales de CoRoT y es además, a diferencia de la función que subyace a los datos de CoRoT, una función analítica, prueba de ello es que la aproximación splines puede ajustarse arbitrariamente bien a este modelo.

Para el modelo analítico-2 con el ruido coloreado añadido a los datos sintéticos calculamos de nuevo las conectividades y la variación del SSE con el número de puntos obteniendo resultados bastante similares a los anteriores (ver Fig. 8.6). Este resultado es coherente ya que añadiendo más complejidad al modelo analítico sólo se puede hacer más difícil de ajustar los datos con la aproximación ARMA(20,1) de modo que las

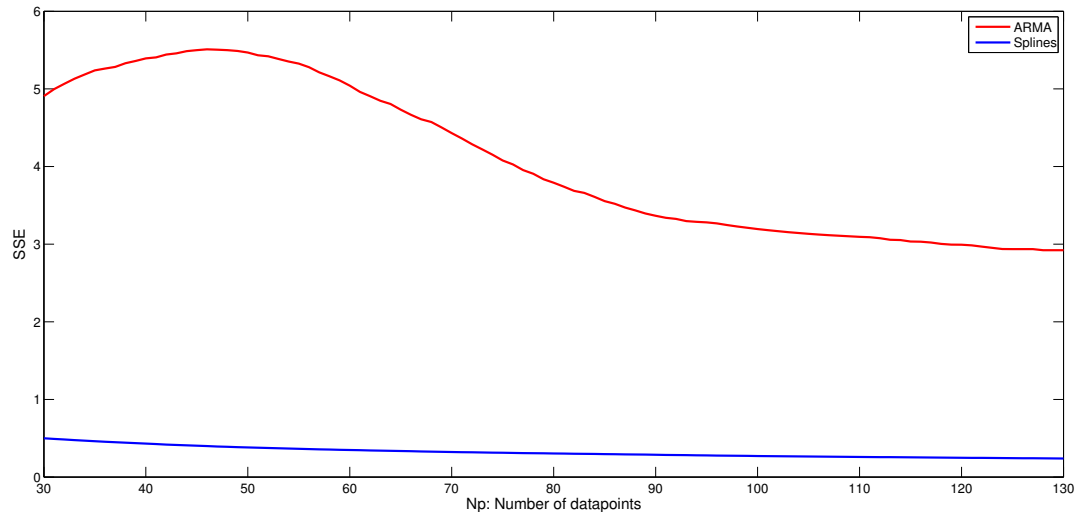


Figura 8.6: Tendencia asintótica del coeficiente SSE con el número de puntos N_p a partir de las conectividades ARMA (en rojo), y las conectividades spline (en azul) para el modelo analítico-2

conectividades y los SSE difieran más aún de las conectividades de las observaciones. No obstante, es interesante observar en la Fig. 8.6 que el ruido coloreado no dificulta el ajuste mediante splines llegándose a un valor muy cercano a cero hasta donde se ha calculado el coeficiente SSE, es decir, $N_p = 130$.

Los resultados obtenidos en esta sección muestran una diferencia significativa entre los datos originales de CoRoTy el modelo analítico. Para profundizar más sobre el origen de la discrepancia encontrada entre modelos y observaciones seguiremos un tratamiento similar con las otras estrellas.

8.2 KIC 006187665

Para descartar posibles efectos instrumentales el objetivo ahora es comparar las conectividades de diferentes series de datos observadas para otras estrellas y con diferentes instrumentos. Por tanto, nos limitamos a partir de ahora al estudio de las conectividades de series de datos reales y no de modelos.

En esta sección hacemos uso de una serie de datos con un muestreo corto (60 segundos) que fue obtenido para la estrella KIC 006187665 por el satélite Kepler (Gilliland et al., 2010) (quarter Q2.2). Esta estrella ha sido clasificada como una híbrida Gdor/DScut en Uytterhoeven et al. (2011), donde además, mediante un análisis de Fourier de la curva de luz se hallaron 659 picos significativos, con 184 de ellos en el rango de las pulsaciones de tipo γ Dor o δ Sct.

Aunque la cantidad de información o contenido en frecuencias de los datos que el modelo ARMA puede ajustar está limitada por sus órdenes, para este caso y para los siguientes que analizaremos aquí usamos los mismos órdenes que para el modelo

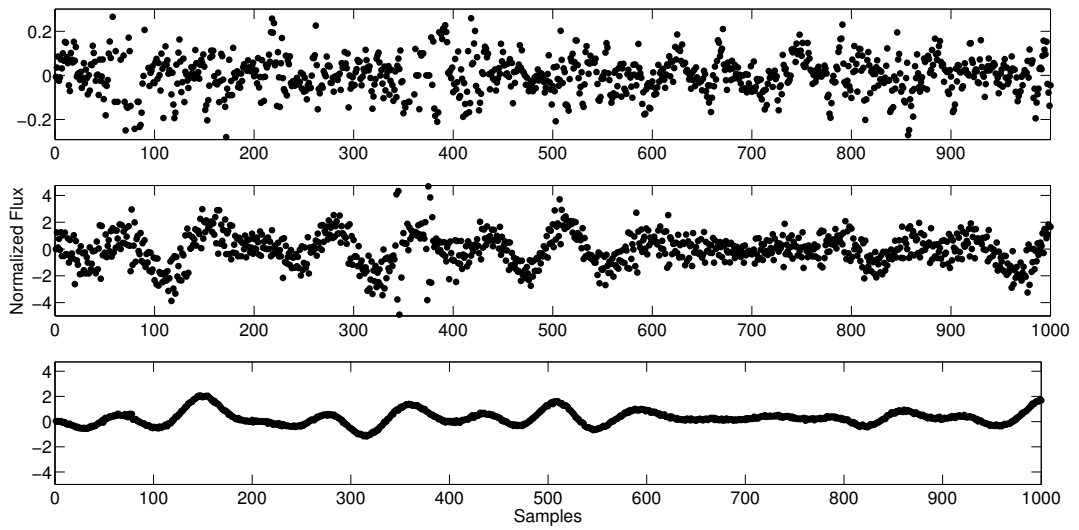


Figura 8.7: Conectividades de la curva de luz de la estrella KIC 006187665. En el panel superior se muestran las conectividades calculadas con la aproximación ARMA, en el panel intermedio se muestran las conectividades spline, y en el panel inferior los datos originales de *Kepler* para su comparación. Nótese las diferentes escalas.

analítico de la estrella HD 174936, es decir, $p = 20$ y $q = 1$, ya que estamos más interesados en las diferencias observadas usando las conectividades que en ajustar el modelo más preciso a los datos. Además, para hacer la comparación más consistente normalizamos estadísticamente todas las series antes de aplicar ninguna de las técnicas comentadas.

La situación que aparece en la Fig. 8.7 es similar a la que encontramos para la estrella HD 174936, donde las conectividades spline están correladas con la señal y las conectividades ARMA muestran una estructura aleatoria casi independiente con una aparente no correlación.

8.2.1 SSE

Nuevamente hemos hallado la tendencia asintótica de la SSE hallada mediante las 2 aproximaciones que utilizamos para calcular las conectividades. Puede apreciarse en la Fig. 8.8 que la tendencia entre $N_p = 30$ y $N_p = 500$ es similar a la observada en el caso anterior salvo pequeñas diferencias. En este caso y en los siguientes, completamos el análisis de conectividades estudiando la representación de estas en el dominio de frecuencias para caracterizar la correlación que surge entre estas y la serie original. Para ello, basándonos en la propia metodología introducida, comprobamos que las conectividades son susceptibles de ser analizadas en el dominio de las frecuencias estudiando primero las conectividades de orden 2, esto es, las conectividades de las conectividades.

8. ANALITICIDAD DE LAS SERIES OBSERVADAS

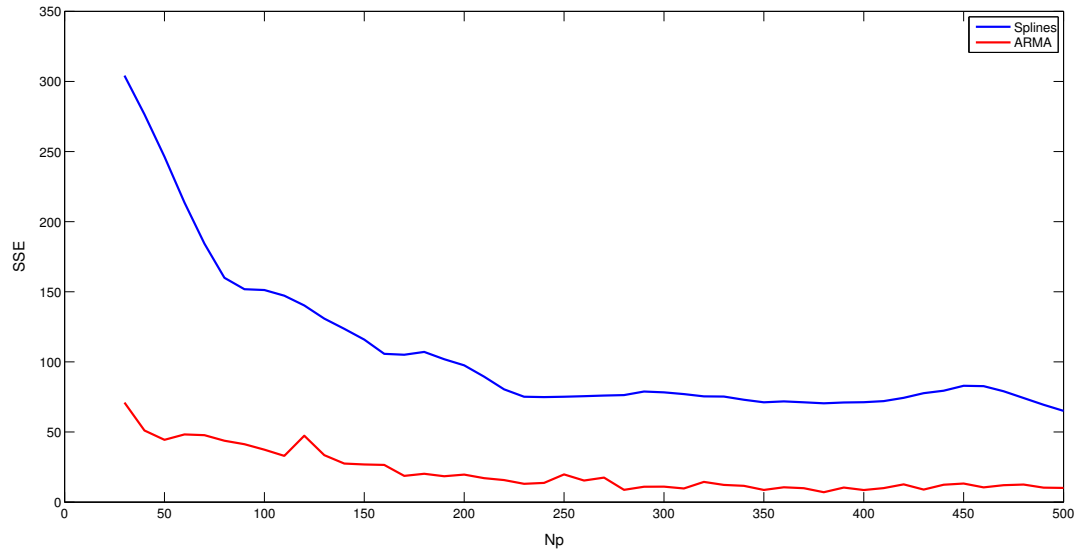


Figura 8.8: Tendencia asintótica del coeficiente SSE con el número de puntos N_p a partir de las conectividades ARMA (en rojo), y las conectividades spline (en azul) para los datos de la estrella KIC 006187665.

8.2.2 Conectividades de orden 2

La analiticidad, entendida como diferenciabilidad infinita, garantiza la convergencia de la expansión en serie de Fourier de una función haciendo posible de ese modo el análisis de Fourier. La prueba de analiticidad aquí descrita indica la diferenciabilidad de la función subyacente a una serie temporal dada. Considerando las conectividades como una nueva serie temporal, la aplicación de esta prueba a éstas nos informa sobre su propia diferenciabilidad. Entonces, si las conectividades de segundo orden (conectividad de la conectividad hallada mediante la aproximación splines) son independientes y con una distribución aleatoria gaussiana, las conectividades de primer orden serán diferenciables. Es legítimo, por tanto, cuando esto ocurre llevar a cabo un análisis de Fourier de las propias conectividades.

Al contrario que las conectividades de primer orden, las de segundo orden (cf. Fig. 8.9) muestran una distribución aleatoria típica de un ruido blanco con distribución gaussiana. Así pues las conectividades de primer orden resultan ser diferenciables y podemos concluir aquí que las técnicas de Fourier pueden ser aplicadas para estudiar la forma de éstas en el espacio de frecuencias.

8.2.3 Análisis espectral de las conectividades

Comparamos ahora el espectro de las conectividades con el de la serie original. Para calcular los espectros utilizamos conjuntos de datos de un tamaño mayor que para el análisis de conectividades para tener una mejor resolución espectral. Así, tomamos

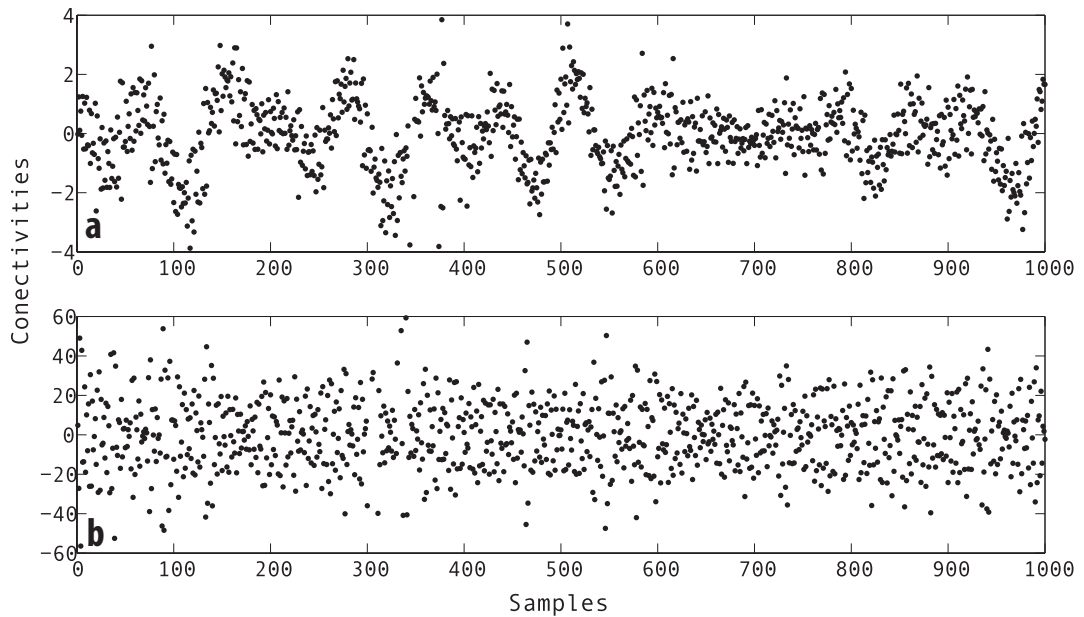


Figura 8.9: Conectividades splines de orden-2 (panel inferior) y conectividades spline de orden-1 (panel superior) para la estrella KIC 006187665.

10000 puntos de la serie original teniendo en cuenta que estos datos no tengan problemas de saltos ni huecos. Usamos unidades arbitrarias para amplitudes y frecuencias ya que, más que la densidad espectral de potencias, nos interesa estudiar las correlaciones y para ello no es necesario usar unidades físicas.

En el caso de las conectividades ARMA (panel superior de la Fig. 8.10), pese a que el espectro no llega a ser completamente plano (como sería de esperar en el caso ideal de un ruido blanco), puede verse que la magnitud de las conectividades es unas 40 veces inferior a la de las halladas con la aproximación de splines. Consideramos que las amplitudes de los picos que aparecen no son significativas y el espectro es aproximadamente el de una serie no correlada.

Por otro lado, vemos que en el caso de las conectividades splines el espectro presenta una estructura en el rango de bajas frecuencias muy similar a la del espectro de la serie original. Además, aparece también una estructura periódica que aumenta de amplitud conforme crecen las frecuencias. Estas periodicidades aparecen a causa de los lóbulos de un alias cercano a la frecuencia de Nyquist introducido por la propia aproximación splines. En efecto, dada la generalización que hemos hecho a la aproximación numérica de derivada que hemos introducido en 7.4, las conectividades pueden expresarse como

$$\mathcal{C}_n = (2\Delta t) \cdot \left[\mathcal{D}_n - \frac{x_{n+1} - x_{n-1}}{2\Delta t} \right] \quad (8.6)$$

donde el cociente que aparece en el lado derecho de la ecuación es la expresión

8. ANALITICIDAD DE LAS SERIES OBSERVADAS

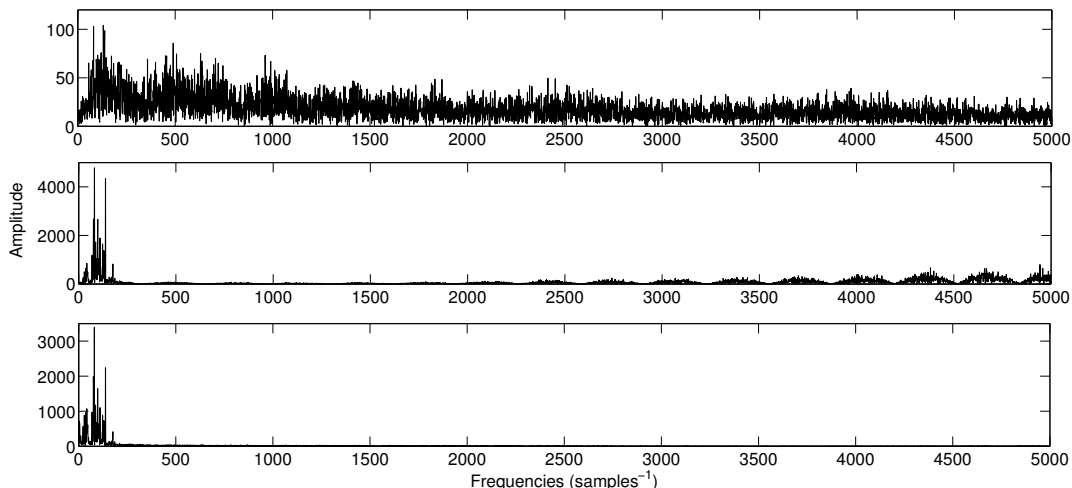


Figura 8.10: Espectro de las conectividades ARMA (panel superior), conectividades spline (panel intermedio) y de la serie original (panel inferior) de la estrella híbrida KIC 006187665.

clásica de la derivada numérica [Burden and Faires \(2011\)](#) y que tiene en cuenta los valores x_{n+1} y x_{n-1} pero no el valor intermedio x_n . Lo mismo es aplicable al cálculo de las conectividades, que por la propia definición tampoco tienen en cuenta la derivada del valor intermedio. Esto, en el caso de la aproximación splines tiene el efecto mencionado del alias cuando las variaciones entre un punto y el siguiente son grandes, ya que en ese caso los splines al ser funciones suaves no ajustan bien la pendiente del punto x_n a partir de los x_{n+1} y x_{n-1} . En el caso de la aproximación ARMA, esto no tiene repercusión sobre el espectro de las conectividades ya que no se restringen al ajuste de funciones suaves y pueden asumir variaciones grandes entre un punto y el siguiente.

Este alias mencionado y sus lóbulos aparecen siempre que hay variaciones significativas entre cada muestreo de la función y en este caso ocurre así pero, como veremos, esto resulta ser el caso también para las demás estrellas estudiadas.

En resumen, las intrigantes propiedades observadas en la estrella HD 174936 aparecen también en esta estrella híbrida observada por *Kepler*, con lo cual descartamos que haya ningún efecto instrumental involucrado en este fenómeno. Teniendo en cuenta que las conectividades están relacionadas con los residuos que resultan del modelado paramétrico, la similitud entre las componentes de bajas frecuencias del espectro de las conectividades spline y el espectro de la serie original puede verse como un fenómeno de autosimilaridad típico de fractales. En este sentido, parece razonable situar el origen de determinados fenómenos que surgen al hacer un *prewhitening* de un espectro como el *plateau* hallado en HD 50844 ([Poretti et al., 2009](#)) o la emergencia de periodos eliminados a distintas escalas en HD 49434 ([Chapellier et al., 2011](#)), en la falta de analiticidad designada aquí como no-conectividad.

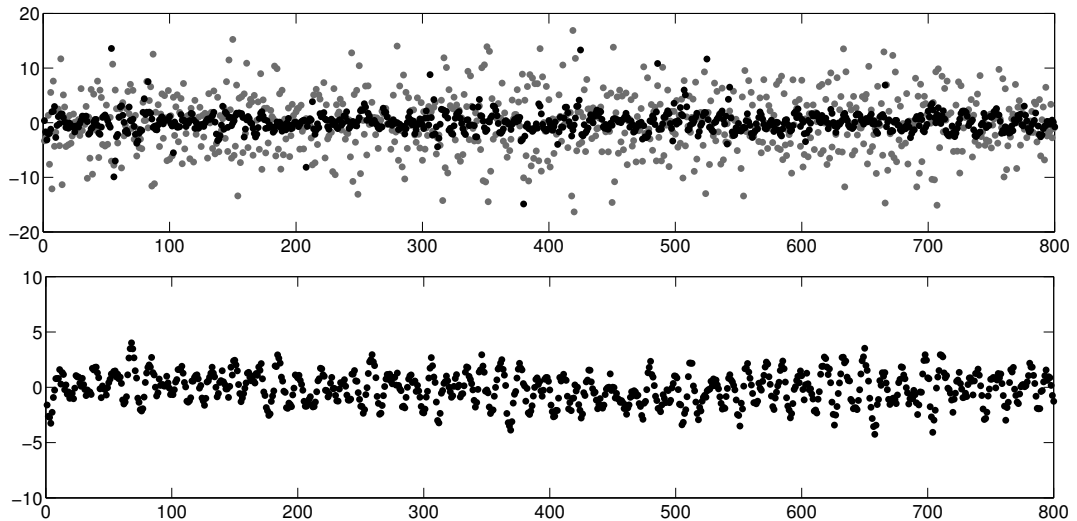


Figura 8.11: Conectividades de la serie de observaciones del Sol. En el panel superior, la aproximación ARMA en negro y la aproximación de splines en gris. En el panel inferior, la serie de datos original de SoHO/GOLF para ser comparada.

8.3 SoHO / GOLF

Los datos usados para esta parte vienen del instrumento GOLF (García et al., 2005) a bordo de la misión SoHO. Este instrumento, GOLF, mide las variaciones de velocidad radial a través del efecto Doppler. Analizar las conectividades de una serie que proviene de otro instrumento con unas características completamente distintas y una física distinta nos permitirá discernir si hay algún efecto relacionado con el tipo de medida que influya en los resultados anteriores.

La Fig. 8.11 muestra los resultados para las conectividades en el panel superior y las velocidades radiales en el panel inferior. Se representan solamente 800 puntos para visualizar mejor las estructuras que aparecen en las series. En el panel superior se muestran las conectividades spline en gris y en negro las conectividades ARMA. La varianza de las conectividades spline es 9.35 veces la varianza de las conectividades ARMA.

Trataremos, a continuación, de ver si el estudio de la tendencia asintótica de la SSE revela más detalle sobre el ajuste.

8.3.1 SSE

En este caso tanto la aproximación spline como la ARMA muestran una tendencia asintótica monótonamente decreciente hasta llegar a $N_p = 1000$, pero la primera tiende a un valor significativamente mayor que cero, mientras que la segunda se aproxima a cero (cf. Fig. 8.12).

8. ANALITICIDAD DE LAS SERIES OBSERVADAS

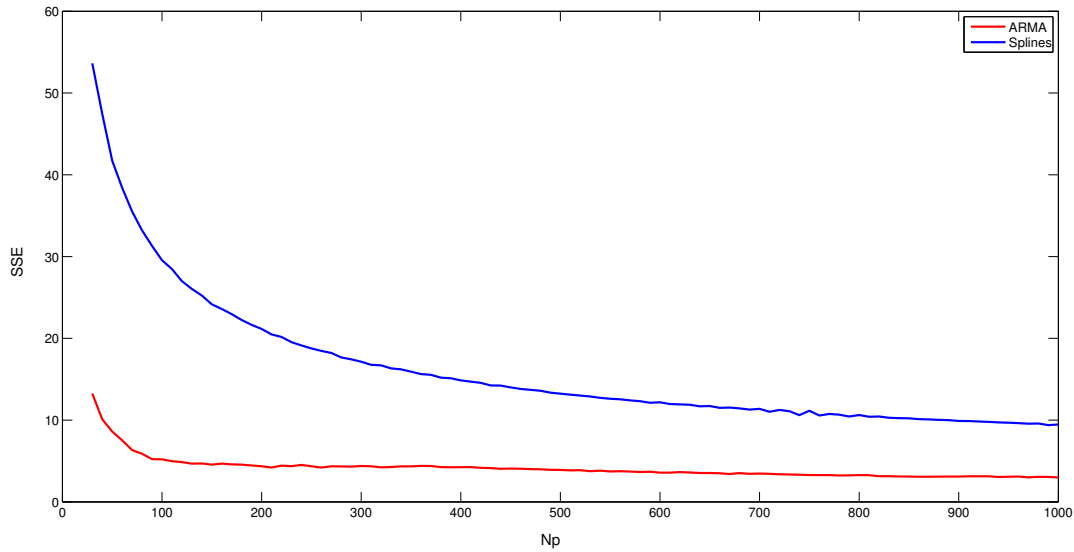


Figura 8.12: Tendencia asintótica del coeficiente SSE con el número de puntos N_p a partir de las conectividades ARMA (en rojo), y las conectividades spline (en azul) para los datos de SoHO/GOLF.

Sin embargo, aún no podemos afirmar que un ajuste de tipo analítico como los splines sea insuficiente para la función que subyace a los datos de SoHO/GOLF. Para obtener una respuesta definitiva hacemos un análisis espectral en el siguiente apartado que nos permitirá investigar las posibles correlaciones entre las series.

8.3.2 Análisis espectral de las conectividades

A pesar de la enorme dispersión hallada en las conectividades splines respecto a las ARMA, no se puede determinar visualmente si hay una estructura residual o si hay algún tipo de correlación entre las conectividades y la serie original (cf. Fig. 8.11). Para comprobar si esto es así hacemos un análisis de Fourier de las conectividades. Puede observarse que el espectro de las conectividades ARMA(20,1) es bastante plano salvo algunas frecuencias significativas que hacen que el espectro difiera del espectro plano de un ruido blanco (Fig. 8.13, panel superior). Esto se debe al orden usado para el modelo ARMA, que limita la precisión del ajuste. El espectro solar requiere más parámetros para capturar todas las características de la serie que en los casos anteriores por ser más complejo, de ahí que este modelo no sea suficiente para un buen ajuste pero sí es adecuado para el estudio comparativo que realizamos en este caso.

Nuevamente se aprecia la misma estructura periódica discutida en el caso de la estrella KIC 006187665. Este es un efecto relacionado con la aproximación utilizada al hacer los cálculos (splines) pero es también un efecto intrínsecamente relacionado con la señal y es, por tanto, revelador sobre el fenómeno físico subyacente. Por otra

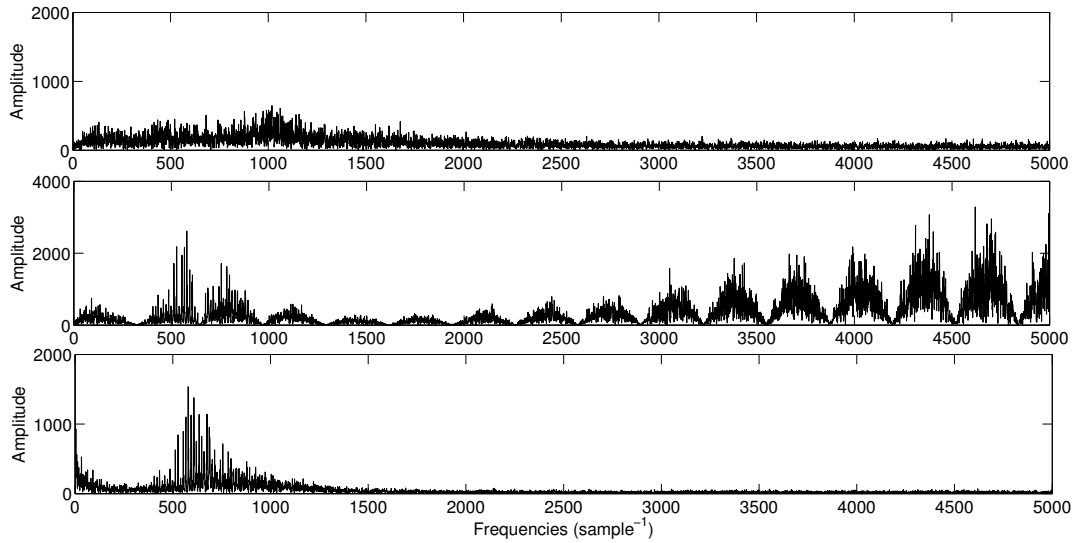


Figura 8.13: Espectro de Fourier de las conectividades ARMA (panel superior), splines (panel medio), y espectro de Fourier de la serie de datos original (panel inferior) de SoHO/GOLF.

parte, aparece también un patrón en frecuencias bajas e intermedias que coincide aproximadamente en posición y forma con el patrón de las oscilaciones de 5 minutos del Sol (ver panel inferior de la Fig. 8.13).

Los resultados que hemos hallado en este y en los anteriores apartados sugieren una estructura fina en la función subyacente de las series procedentes de CoRoT, *Kepler*, y SoHO/GOLF que hace que esta sea no analítica. Hemos comprobado que se obtienen resultados similares en 3 casos de series obtenidas por instrumentos distintos, originadas por observaciones de objetos distintos, y con técnicas diferentes. Esta aparente ubicuidad del fenómeno en estrellas pulsantes se contrastará en el siguiente apartado mediante el estudio de una estrella constante.

8.4 HD 181072

Esta es una estrella de tipo espectral A2 que fue observada durante el primer *Long Run* de observaciones de CoRoT y al estudiar la variabilidad fotométrica en el seismocampo de las cámaras de CoRoT se ha clasificado como constante. La curva de luz de esta estrella presenta solamente variaciones aleatorias con una distribución gaussiana debido al ruido fotonico. Este es un caso ideal para calibrar el método de las conectividades ya que al no haber otras variaciones en la curva de luz la función subyacente debe ser una función lineal con pendiente constante y las conectividades deben tener el mismo tipo de distribución aleatoria que la serie original.

En la Fig. 8.14 se puede apreciar que las 3 series - conectividades ARMA, conectividades spline y la serie original - presentan distribuciones similares. Aunque las

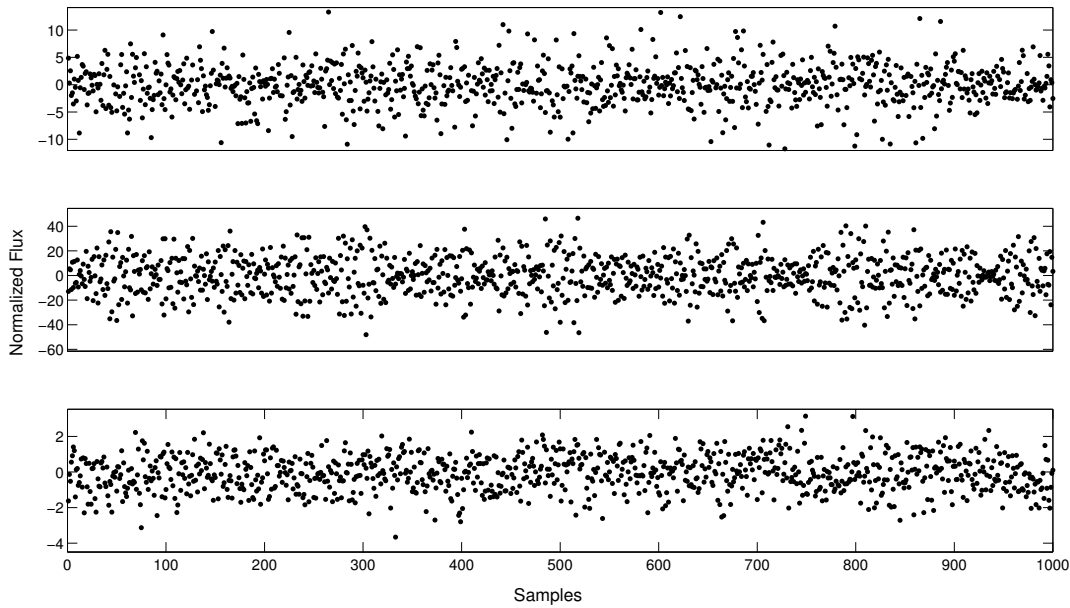


Figura 8.14: Conectividades de la serie de observaciones de la estrella HD 181072 por el satélite CoRoT. En el panel superior se muestra la aproximación ARMA, en el panel medio la aproximación de splines, y en el panel inferior, la serie de datos original para ser comparada.

magnitudes de las conectividades son grandes esto no supone un problema siempre que estas se puedan asimilar a unos residuos aleatorios e independientes con distribución gaussiana como es el caso.

Evaluaremos, a continuación, la tendencia asintótica del SSE para comprobar que las 2 aproximaciones tienen un comportamiento coherente con lo establecido en el anterior párrafo.

8.4.1 SSE

En la Fig. 8.15 puede apreciarse como, tanto la SSE de las conectividades ARMA como la de las conectividades splines son funciones monótonamente decrecientes, lo que corrobora las conclusiones del apartado anterior.

Las dos gráficas anteriores parecen mostrar que las conectividades de la curva de luz de la estrella HD 181072 tienen un comportamiento propio de una función analítica. No obstante, el análisis espectral de las conectividades nos dará una prueba definitiva sobre la no-correlación de los resultados.

8.4.2 Análisis espectral de las conectividades

Al comparar el espectro de las conectividades ARMA (Fig. 8.16, panel superior) con el espectro de la serie original (panel inferior) se observa que ambos tienen la misma distribución en frecuencias, es decir, un espectro plano propio de un ruido no correlado.

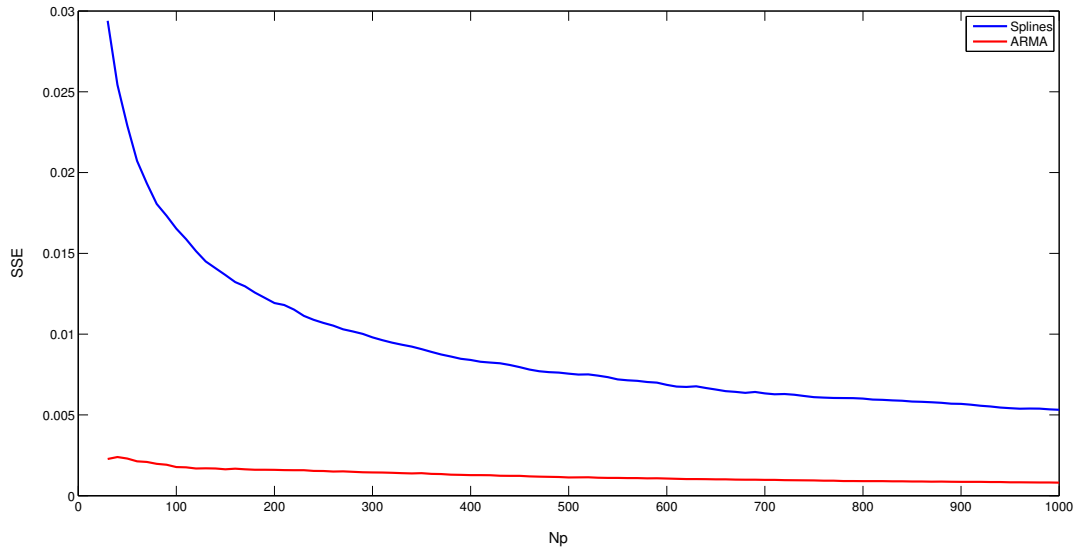


Figura 8.15: Tendencia asintótica del coeficiente SSE con el número de puntos N_p a partir de las conectividades ARMA (en rojo), y las conectividades spline (en azul) para los datos de la estrella HD 181072 observada por CoRoT.

En cambio, el espectro de las conectividades splines (panel medio) muestra un patrón que diferencia éste de los demás. No obstante, a frecuencias bajas el espectro resulta ser completamente plano y el patrón que domina las altas frecuencias resulta ser nuevamente una modulación debido a la ventana espectral que introduce la aproximación splines para series con una variabilidad alta entre muestreos. Este fenómeno tiene una explicación clara en este caso dada la alta dispersión del ruido fotónico. Sin embargo, la ubicuidad de este fenómeno puede revelar propiedades de las curvas de luz no estudiadas anteriormente.

Obviando este efecto que domina las altas frecuencias el espectro que resulta de las conectividades splines es plano también y podemos concluir que la función que subyace a la curva de luz de la estrella HD 181072 observada por CoRoT es una función analítica cuya derivada puede ser calculada numéricamente a través de su definición usual de acuerdo a la definición dada en Eq. 7.4.

Los resultados mostrados en esta sección contrastan con los anteriores y han servido para calibrar el método mostrando que con éste podemos averiguar si la función que subyace a una serie temporal es analítica o no y en definitiva, con ello determinar la validez del desarrollo en series de Fourier.

8.5 Discusión

Hemos establecido en esta parte de la memoria como criterio matemático que una serie temporal debe satisfacer para que el análisis de Fourier de la serie sea consistente su analiticidad. Para ello hemos desarrollado un método basado en el estudio de

8. ANALITICIDAD DE LAS SERIES OBSERVADAS

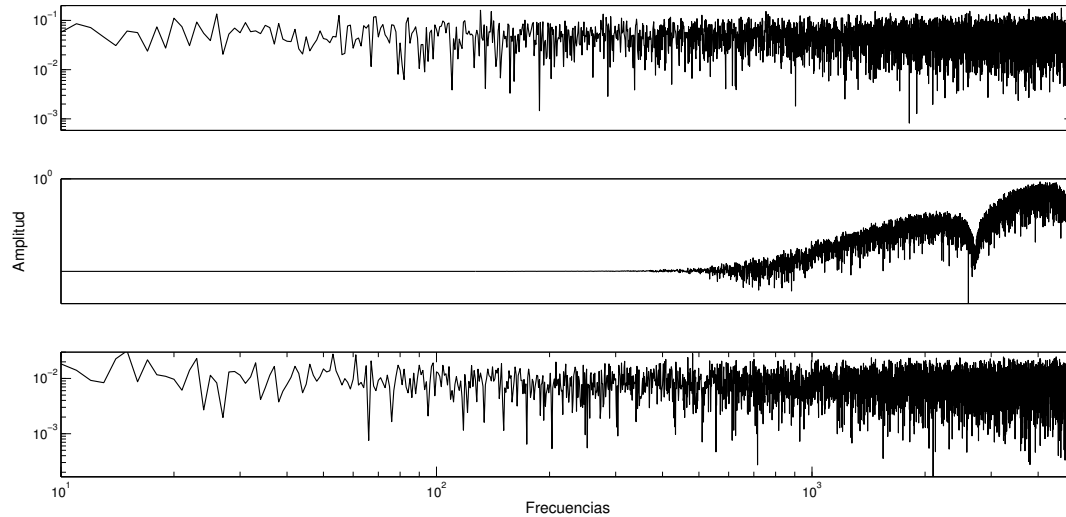


Figura 8.16: Espectro de Fourier de las conectividades ARMA (panel superior), splines (panel medio), y espectro de Fourier de la serie de datos original (panel inferior) de la estrella constante HD 181072 observada por CoRoT

la diferenciabilidad de la función subyacente por medio de las conectividades entre puntos.

La definición de las conectividades dada a través de la ecuación 7.4 generaliza el concepto de derivabilidad de modo que la definición usual resulta como caso particular cuando la función subyacente es diferenciable. En ese caso, las conectividades se anulan por completo. Con esta definición se puede considerar que la función es diferenciable también en el caso más corriente en que las conectividades tengan una distribución propia de un ruido no correlado, independiente y con distribución gaussiana.

Tanto para la estrella HD 174936 como para KIC006187665 y la serie de SoHO/GOLF los resultados que muestran el estudio asintótico de la SSE y el análisis espectral de las conectividades indican que la función subyacente a estas series discretas no es analítica y por lo tanto, la convergencia de su expansión en series de Fourier no está garantizada. Las correlaciones entre las conectividades spline y las correspondientes series originales son evidentes por las figuras, pero hemos calculado además, los coeficientes de correlación de Pearson para descartar la hipótesis nula (no correlación).

Hallamos también los valores de probabilidad para el test de la hipótesis nula usando una distribución t-Student (ver tabla 8.2). Considerando los valores despreciables obtenidos para P_s , la hipótesis nula debe ser descartada bajo cualquier límite de validez computable. Por otro lado, las conectividades ARMA pasan el test con la elección estándar del límite en $P < 0.05$ (i.e. sólo en el 5% de los casos las conectividades ARMA estarían correladas con la serie original). Por tanto, podemos decir que

Tabla 8.2: Coeficientes de correlación de Pearson (ρ) y las probabilidades de no correlación (P) asociadas a las conectividades ARMA y splines de las series temporales estudiadas en este trabajo. Los subíndices 'a' y 's' se corresponden a las aproximaciones ARMA y spline, respectivamente.

Serie Temporal	ρ_a	P_a	ρ_s	P_s
CoRoT 7613	-0.0075	0.8139	0.4243	5.8247e-45
Modelo analítico	0.0072	0.8199	8.371e-4	0.9789
KIC 006187655	-0.0257	0.4176	0.5553	5.8039e-82
SoHO/GOLF	-0.0692	0.0903	0.1805	8.6429e-6

los modelos ARMA capturan la componente no analítica de la señal.

Podemos descartar efectos instrumentales en los resultados dado que se han analizado series de datos procedentes de distintos instrumentos basados en técnicas distintas (fotometría y velocidades radiales).

En conclusión, las pruebas que hemos realizado muestran que la función subyacente a las series temporales de estrellas pulsantes estudiadas no es analítica y por lo tanto la descripción basada en frecuencias de Fourier puede ser incorrecta.

Terminamos ahora este capítulo conectando los resultados con las cuestiones planteadas inicialmente en el Cap. 6 en torno a la teoría de distribuciones.

En la teoría del análisis espectral de series temporales se muestra que un proceso gaussiano de cuadrado sumable puede ser representado siempre por un modelo lineal (Priestley, 1988). Si esta condición se mantiene, la expansión en serie en el dominio de las frecuencias es convergente y de este modo el periodograma es un estimador consistente y no sesgado de la función densidad espectral de la señal. Cuando esto no ocurre, como es el caso de las series temporales estudiadas en este capítulo, el periodograma es un estimador inconsistente del espectro de la señal. Por otro lado, dado que la serie no puede ser representada por un modelo lineal su distribución no puede ser normal, ya que por definición cada combinación lineal de una distribución normal debe ser necesariamente normal también. Entonces, las curvas de luz estudiadas tienen una distribución no gaussiana y además se demuestra que sus series discretas vienen de una función no analítica, y de ahí que no sean series de cuadrado sumable.

8. ANALITICIDAD DE LAS SERIES OBSERVADAS

Parte IV

Conclusiones y trabajo futuro

Conclusiones

En este capítulo resumimos las conclusiones principales de este trabajo de tesis y las consecuencias que se derivan de ellas. Estas son:

- Se ha desarrollado una nueva técnica para el relleno de huecos basada en modelos ARMA orientada a la sustitución de los datos no-válidos de las curvas de luz de CoRoT. El método consiste en el modelado local de los segmentos de datos válidos en torno a cada hueco y una predicción a partir de estos modelos de los datos en el hueco. Este método supera las limitaciones intrínsecas al método estándar actualmente utilizado para corregir las curvas de CoRoT y que se basa en una interpolación lineal. El código MIARMA desarrollado en este trabajo ha sido recomendado por el Comité Científico de CoRoT para ser el algoritmo que corrija los datos marcados como no válidos por el satélite en el Nivel 3.
- Se ha desarrollado una rutina de detección de frecuencias (FRECS) para los tests del método de interpolación que no requiere ajustes de mínimos cuadrados ni técnicas de *prewhitening*, de modo que no introduce ningún sesgo debido al error en la estimación de los parámetros.
- Los resultados obtenidos al aplicar el método de relleno de huecos a la curva de luz de la estrella de tipo solar HD 49933 muestran que, si bien los efectos de la interpolación sobre el espectro son menores en series con el contenido en frecuencias que poseen estrellas de este tipo, no son despreciables y deben ser tenidos en cuenta. En concreto, se muestra que la aplicación del método reduce los efectos causados por las interpolaciones de tipo analítico, haciendo posible una mejor determinación de los parámetros de cada línea.

9. CONCLUSIONES

- El análisis de periodicidades de las frecuencias detectadas en la curva de luz de HD 174966 muestra que la interpolación ARMA permite una mejor determinación de las cuasi-periodicidades de las frecuencias detectadas en estrellas δ Sct.
- El hecho de que el espectro de potencias de una estrella Be con pulsaciones de bajas frecuencias se vea mucho más afectado por los alias que en el test numérico llevado a cabo en § 5.1.1.1 puede implicar que la señal contenida en su curva de luz no provenga de una función subyacente de tipo analítico.
- La interpolación ARMA en la curva de luz de la estrella tipo solar HD 49933 revela una estructura fina que no se detecta mediante análisis de Fourier y que, por la propia definición matemática de los modelos ARMA, no puede identificarse con ningún tipo de ruido. Esta estructura es una propiedad inédita en cualquier tipo de series temporales de estrellas pulsantes.
- Los resultados obtenidos en la parte II de esta tesis sugieren que al desajuste entre modelos teóricos y observaciones en astrosismología puede contribuir el sesgo en las herramientas de análisis que se utilizan. Para profundizar en este aspecto hemos diseñado un test de analiticidad de la función subyacente a una serie temporal. El test está basado en la definición de un nuevo parámetro que hemos llamado conectividad, que está estrechamente relacionado al coeficiente de no-diferenciabilidad introducido por Wiener.
- Hemos utilizado este test para caracterizar las propiedades de la función subyacente de series temporales procedentes de observaciones por satélite de estrellas pulsantes. En concreto, hemos caracterizado las series de la δ Sct HD 174936 observada por CoRoT, de la estrella KIC 006187665 observada por *Kepler*, de velocidades radiales del Sol observadas por SoHO/GOLF, y de la estrella constante HD 181072 observada por CoRoT. Definitivamente, los resultados muestran que, excepto la estrella constante, las series temporales de las demás estrellas provienen de una función no analítica.
- La no analiticidad hallada en las series temporales estudiadas en la Parte III implica que la función subyacente cuyo muestreo da lugar a las series observadas puede no ser desarrollable en una serie de Fourier.
- El efecto puesto en evidencia en la Parte III del trabajo de tesis es independiente de la técnica, instrumento, y objeto observado. Se ha demostrado que la no analiticidad es una propiedad intrínseca de las series temporales analizadas.

-
- El periodograma de las series temporales, obtenido a través de transformadas discretas de Fourier, que debe converger a un valor real, sin embargo, no está garantizado que sea una estimación no sesgada de la densidad espectral de la función subyacente.
 - En consecuencia, los resultados apuntan a que el concepto de detección no ambigua de una frecuencia, identificada en astrosismología como un máximo estadísticamente significativo en el periodograma, debería ser reconsiderado.

9. CONCLUSIONES

Capítulo 10

Trabajo futuro

Los resultados obtenidos muestran que el uso de herramientas tradicionales para la determinación de frecuencias en series temporales de estrellas pulsantes conlleva sesgos de interpretación cuando se comprueba que no se verifican las hipótesis en las que están basadas los teoremas de base de dichas herramientas. Consecuentemente, se han desarrollado algoritmos para comprobar estas hipótesis y llevar a cabo un análisis de frecuencias no sesgado. Las líneas de trabajo futuro pasan por el aprovechamiento de dichas herramientas y la exploración de las nuevas perspectivas que se abren por su aplicación masiva a las bases de datos ya existentes. Concretamente, el test de analiticidad que comprueba la consistencia del análisis armónico de las series temporales de estrellas pulsantes haría imperativo buscar una descripción alternativa de sus curvas de luz. En este sentido, las líneas de trabajo futuras son:

- Utilizar las conectividades para corregir las series temporales de la no analiticidad generando así nuevas series que puedan ser descritas en términos de frecuencias de Fourier.
- Aplicación del método de las conectividades a otras estrellas distintas a la HD 174936, que no tengan envoltura convectiva.
- Profundizar en el estudio de la escala de la no analiticidad descubierta en las funciones subyacentes a las series temporales de estrellas pulsantes.
- Buscar una interpretación física a las frecuencias detectadas, lo que nos llevaría a estudiar la relación de la no analiticidad con las observaciones y su posible origen en el concepto de teoría de la medida.
- Estudiar la relación de la no analiticidad con los distintos problemas que surgen al ajustar modelos y observaciones como, por ejemplo, el enorme número de frecuencias detectadas y su rango entre las estrellas de tipo δ Sct, así como las

10. TRABAJO FUTURO

diferencias conocidas como efectos de superficie entre frecuencias observadas y predichas en estrellas de tipo solar.

Apéndices

El programa de análisis de frecuencias FRECS

Para la detección de frecuencias se ha desarrollado un algoritmo de detección de picos basado en el periodograma Lomb-Scargle y asumiendo ruido blanco, es decir, toda componente sea estocástica o no, cuyos valores no son estadísticamente independientes y con distribución gaussiana se toma como parte de la señal cuyas frecuencias son analizadas.

Una vez reconstruida la señal en los huecos, la serie que recibe el programa FRECS como input está muestreada uniformemente y por tanto, puede ser estudiada mediante una FFT (Fast Fourier Transform). Esto hace el algoritmo computacionalmente mucho más eficiente que aquellos que requieren un nuevo ajuste cada vez que se detecta una frecuencia. En lugar de realizar un prewhitening para eliminar los efectos del leakage, el programa FRECS hace uso de una función ventana de tipo Taylor cuya respuesta en frecuencias puede verse en la Fig. A.1 en azul. Teniendo en cuenta la respuesta en frecuencias de la ventana Taylor, los lóbulos que aparecen en torno a cada pico en el espectro de frecuencias de la serie temporal debido a la convolución de la función sinc (transformada de la función caja que representa el intervalo de observación en el dominio del tiempo) se ven atenuados por completo y puede llevarse a cabo una detección de frecuencias pico a pico en base a criterios de significancia simplemente sin tener que hacer uso de ningún método adicional para eliminar contribuciones espúreas. En la literatura existen numerosas definiciones de funciones ventana, especialmente en la literatura de ingeniería y análisis de señal. La respuesta en frecuencias que da cada ventana varía mucho según la forma de la función en el dominio del tiempo. Así, una ventana de tipo Hamming tiende a concentrar la potencia en el pico principal con lo que éste tiene una anchura muy estrecha a costa de que los lóbulos en torno al pico se reduzcan poco. Por contra, una ventana como la Flat Top reduce considerablemente los lóbulos pero se pierde resolución ya que el pico

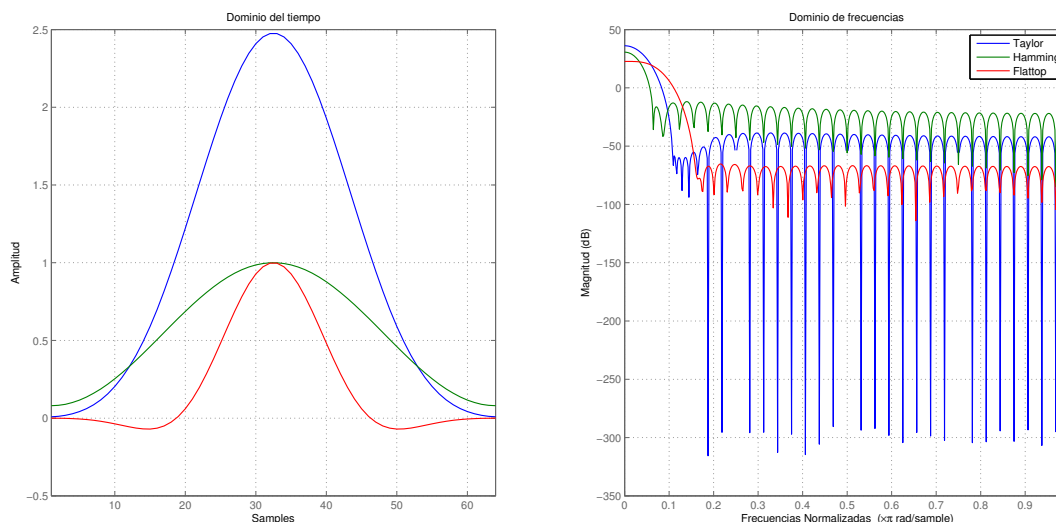


Figura A.1: La ventana Taylor en el dominio del tiempo y su respuesta en frecuencias (azul) comparada con una ventana Hamming (verde) y una Flatop (rojo). Factor de leakage de 0%, atenuación relativa de los lóbulos de -74.8 dB, anchura del pico a -3 dB igual a 0.050781.

principal aumenta en anchura. La elección correcta de la función ventana depende de las características de la serie a analizar. Así, si la observación es corta y la resolución espectral, que es inversamente proporcional, no es menor que la separación entre las frecuencias de las componentes de la señal, entonces es mejor una ventana que tenga una anchura del pico central reducida como la Hamming. Sin embargo, para el caso de los datos de CoRoT en que la resolución espectral puede ser del orden de $0.1 \mu\text{Hz}$, es mejor elección una ventana en la que los lóbulos se reduzcan más para no introducir frecuencias espúreas en la detección.

La ventana Taylor es una función paramétrica que se puede elegir de forma que se obtenga una reducción de los lóbulos sin una considerable pérdida de resolución. Concretamente, en este caso se han fijado los parámetros de la ventana de forma que se obtenga un decaimiento relativo entre el pico principal y los 4 primeros lóbulos de -100 dB. Esto garantiza que los principales alias de cada frecuencia, que son los de mayor amplitud, queden atenuados por completo.

El límite de detección que usa FRECS para decidir si una frecuencia es significativa o no está basado en el cociente señal-ruido $\text{SNR} > 4$ sugerido en [Breger et al. \(1993\)](#). Este es un criterio empírico que ha sido ampliamente utilizado en la literatura. La estimación del error de detección se realiza mediante la aproximación de Montgomery y O'Donoghue ([1999](#)).

El rango explorado por defecto es de 0 a $1000 \mu\text{Hz}$ pero el programa admite un modo manual que permite explorar un rango distinto.

Para obtener las frecuencias de pulsación es importante que en la serie temporal de

entrada del programa FRECS se eliminan posibles trends para quitar la contribución de las frecuencias muy bajas que puede distorsionar el espectro obtenido a partir del cual se realiza la detección.

```

function [F, pfftSmil, f] = frecs(S, dt, varargin)
% Function [F, pfftSmil, f] = frecs(S, dt, varargin) detect significant
% frequencies in a time series by identifying local maxima in the
% periodogram.
% Input: S = time series in magnitudes
% dt = time sampling
% [F, pfftSmil, f] = frecs(S, dt, 'manual') - input parameters
% manually
% Output: F - a matrix of frequencies sorted by amplitude with the
% following columns:
% 1) Number
% 2) Frequency
% 3) Frequency Error
% 4) Amplitude in mmag
% 5) Amplitude Error
% 6) Phase angle
% 7) Phase Error
% 8) Signal-to-Noise Ratio (SNR)
% pfftSmil - periodogram in the range selected
% f - frequencies for the periodogram

% Notes: It works only for band-limited signals and a time sampling
% such
% that Nyquist is much higher than the highest signal frequency.
%
% Description: The program performs the FFT of the time series
% windowed
% with a specified window function. The FFT is padded with a factor rx
% and the resulting transform is used to calculate the periodogram.
% The
% noise level is obtained averaging the bins on a predefined range of
% frequencies. Then, significant frequencies are obtained by peak
% detection above a signal-to-noise ratio limit. The output of the
% program contains frequencies, amplitudes, phases, and their
% corresponding errors calculated using Montgomery and O'Donoghue
% approximation. Additionally, the SNR of each peak is given.
%
% Author(s): Javier Pascual-Granado
% $Date: 19/02/2014 19:55$

disp('FREQUENCY_DETECTION_ALGORITHM');
fprintf('_by_Pascual-Granado, _J.\n\n');

L = length(S);
S = reshape(S, L, 1); % convierte en columna
fprintf('Data must be in magnitudes\n');
fprintf('The sampling must be in seconds\n\n');

% NaN and Inf data is fixed
if (isempty(isnan(S))==0),
    S(isnan(S))=mean(S(isnan(S))==0); % sustituimos los NaN
end

```


A. EL PROGRAMA DE ANÁLISIS DE FRECUENCIAS FRECS

```

if (isempty(isinf(S))==0),
    S(isinf(S))=mean(S(isinf(S)==0)); % sustituimos los Inf
end

%% Parameters
i = find(strcmp(varargin, 'manual'), 1);
if isempty(i)==0,
    q = input('Detection_limit_(SNR): ');
    rx = input('Oversampling: ');
    Nini = input('Noise_range_(initial): ');
    Nend = input('Noise_range_(end): ');
    u = input('Freqs._in_c/d_1)_c/d_2)_microhz_3)_no_units: ');
    freqm = input('Max._frequency_to_explore: ');
else
    fprintf('<<Default_values>>\n');
    q = 4; % l mite de detecci n (default)
    fprintf('_Detection_limit_(SNR):_%d\n', q);
    rx = 20; % Oversampling (default)
    fprintf('_Oversampling:_%d\n', rx);
    freqm = 1000.0;
    fprintf('_Max._frequency_to_explore:_%8.3f\n', freqm);
    Nini = 0;
    Nend = 2000; % rango para hallar el ruido
    fprintf('_Noise_range:_%8.3f_to_%8.3f\n', Nini, Nend);
    fprintf('Frecuencias_are_in_microhertz\n');
    u = 2; % unidades
end

if u == 1,
    dt = dt/86400; % sampling in days
    fnyq = 1/dt/2;
    fprintf('_Nyquist_frequency:_%8.4f_d^-1\n', fnyq);
    disp('Frecuencias_in_c/d')
    fprintf('_Oversampling:_%d\n', rx);
    T = L*dt; % tiempo de observacion
    fR = 1/T; % Rayleigh resolution
    df = fR/rx; % paso en frecuencias
    f = (0:df:freqm); % vector de frecuencias en d^-1
    fprintf('_Rayleigh_frequency:_%8.4f_d^-1\n', fR);
    fprintf('_Frequency_step:_%8.4f_d^-1\n', df);
    disp('Amplitudes_in_mmag')
    cstf = sqrt(3)/pi/T;
elseif u == 2,
    fnyq = 1e6/dt/2;
    fprintf('_Nyquist_frequency:_%8.3f_muHz\n', fnyq);
    T = L*dt; % tiempo de observacion
    fR = 1e6/T; % Rayleigh resolution
    df = fR/rx; % paso en frecuencias
    f = (0:df:freqm); % vector de frecuencias en muHz
    fprintf('_Rayleigh_frequency:_%8.3f_muHz\n', fR);
    fprintf('_Frequency_step:_%8.3f_muHz\n', df);
    disp('Amplitudes_in_mmag')
    cstf = 1e6*sqrt(3)/pi/T;
else
    fnyq = 1/dt/2;
    fprintf('_Nyquist_frequency:_%8.3f\n', fnyq);

```

```

    T = L*dt; % tiempo de observacion
    fR = 1/T; % Rayleigh resolution
    df = fR/rx; % paso en frecuencias
    f = (0:df:freqm); % vector de frecuencias en uds inversas
    fprintf(' Rayleigh frequency : %8.3f\n',fR);
    fprintf(' Frequency step : %8.3\n', df);
    disp(' Amplitudes_in_mmag')
    cstf = sqrt(3)/pi/T;
end

%% Main code

N = rx*L;

% Noise range indexes
Nini = 1+round(Nini/df);
Nend = round(Nend/df);

% Windowing with Taylor function. Final amplitudes are calculated from
  S1
% (unwindowed time series)
S1 = S;
w = taylorwin(L,6,-100);
S = w.*S;

% FFT with rx padding
if rx==1,
    fftS1 = fft(S1);
    fftS = fft(S);
else
    fftS1 = fft(S1,N);
    fftS = fft(S,N);
end

% Periodogram
pfftS1 = 2*abs(fftS1)/L;
pfftS = 2*abs(fftS)/L;

% Frequency range explored
pfftSmil1 = pfftS1(1:length(f));
pfftSmil = pfftS(1:length(f));

% Minimum peak separation
if rx>2,
    sep = rx-1;
else sep = 2;
end

% To avoid problems with power close to zero frequency we do not count
  the
% first i0 bins
i0 = 2;
pfftSmil1(1:i0)=0;
pfftSmil(1:i0)=0;

% Noise level estimation

```

A. EL PROGRAMA DE ANÁLISIS DE FRECUENCIAS FRECS

```

noise = mean(pfftS(Nini:Nend)); % noise level

% Finding local maxima
[pks,locs] = findpeaks(pfftSmil,'minpeakdistance',sep);

% Original amplitudes are recovered
pks1 = pfftSmil1(locs);

% Significant frequencies are determined using factor q as limit for SNR
ind = find(pks > q*noise);
pks1 = pks1(ind)';

if isempty(ind)==1,
    fprintf('<<No significant frequencies detected>>\n');
    return;
else
    locs1 = locs(ind);

    if u==1 || u==2,
        % amplitudes and noise changed into mmag
        pks1 = 1e3*pks1;
        noise = 1e3*noise;
        % significant frequencies
        freq = f(locs1);
        % initial phase of the significant frequencies
        fase = angle(fftS(locs1));
    else
        freq = f(locs1);
        pks1 = pks(ind)';
        fase = angle(fftS(locs1));
    end

    const = sqrt(2/pi);

    % amplitud errors in mmag
    errA(1:length(pks1)) = const*noise;
    % phase errors
    errfi = const*(noise./pks1);
    % frequency errors
    errfreq = errfi*cstf;

    % Signal-to-Noise Ratio
    snr = pks1/noise;
    lf = length(freq);
    nnn = 1:lf;

    freq = reshape(freq,lf,1);
    errfreq = reshape(errfreq,lf,1);
    pks1 = reshape(pks1,lf,1);
    errA = reshape(errA,lf,1);
    fase = reshape(fase,lf,1);
    errfi = reshape(errfi,lf,1);
    snr = reshape(snr,lf,1);

    F0 = [freq errfreq pks1 errA fase errfi snr];
    % sort by amplitude

```

```
F0 = sortrows(F0,3);  
% Matrix containing the output of the program  
F = [nnn' flipud(F0)];  
end
```

A. EL PROGRAMA DE ANÁLISIS DE FRECUENCIAS FRECS

Índice de figuras

2.1	Diagrama de contornos de la parte real de los armónicos esféricos. Los contornos positivos se indican por líneas continuas y los contornos negativos por líneas punteadas. El eje de $\theta = 0$ se ha inclinado 45° hacia el observador y viene indicado por un punto. El ecuador se muestra mediante “++++”. Los casos ilustrados se corresponden a: a) $\ell = 1, m = 0$; b) $\ell = 1, m = 1$; c) $\ell = 2, m = 0$; d) $\ell = 2, m = 1$; e) $\ell = 2, m = 2$; f) $\ell = 3, m = 0$; g) $\ell = 3, m = 1$; h) $\ell = 3, m = 2$; i) $\ell = 3, m = 3$; j) $\ell = 5, m = 5$; k) $\ell = 10, m = 5$; l) $\ell = 10, m = 10$. (figura obtenida de Christensen-Dalsgaard 2003)	13
2.2	Curva de luz de la estrella HD 49434 y residuos tras eliminar distintos conjuntos de frecuencias ajustadas. (figura obtenida de Chapellier et al. 2011)	15
3.1	Ventana espectral de una serie de velocidades radiales tomadas desde tierra. Nótese los lóbulos debido al periodo de observación. La serie está constituida por 57 medidas tomadas desde el observatorio de Calar Alto durante una noche de observación. El espectro mostrado se ha sobremuestreado 20 veces para mostrar el detalle de los lóbulos alrededor del pico en $\nu = 0$	33
3.2	Ventana espectral típica en una curva de luz de <i>Kepler</i> . La serie consta de 41443 puntos con un muestreo de 60 s. aproximadamente haciendo un total de 30.79 días observados. Nótese que los alias aparecen con amplitudes muy bajas.	34
3.3	Ventana espectral de la observación de la estrella HD 50890 que fue observada durante el Initial Run (IRa01) de CoRoT durante 56.7 días. Nótese los alias a las frecuencias de 2 cd^{-1} , a 13.97 cd^{-1} y múltiplos de estas.	35

ÍNDICE DE FIGURAS

5.1	Ventana utilizada para generar los huecos en los tests a partir del status de los datos observados por CoRoT en el LRc02. En ordenadas se representan los valores del status de los datos, que toman valor cero cuando no se ha producido ningún evento durante la observación. Nótese la periodicidad introducida por los pasos a través de la Anomalia Suratlántica.	59
5.2	Realización de un modelo analítico compuesto por una señal de bajas frecuencias y ruido blanco gaussiano.	61
5.3	Estimación del orden ARMA(p,q) para el modelo que se utiliza en la reconstrucción de los huecos de la serie sintética compuesta por una señal de bajas frecuencias y ruido blanco gaussiano. Nótese como se alcanzan mínimos en (8,4) y (4,8).	62
5.4	Comparación entre el método ARMA e interpolación lineal para el relleno de huecos. En el panel superior se muestra un segmento de 0.75 días de la serie generada con el modelo analítico de bajas frecuencias. Los segmentos marcados en rojo son los que han sido interpolados en el panel central (interpolación lineal) y en el inferior (interpolación ARMA).	64
5.5	Periodograma en escala logarítmica de la serie interpolada linealmente (en verde) comparada con el periodograma de la serie con huecos (azul).	65
5.6	Periodograma en escala logarítmica de la serie interpolada con ARMA (en rojo) comparada con el periodograma de la serie con huecos (azul). Nótese que en este caso las amplitudes de los frecuencias principales de la señal son mayores que en la figura anterior.	65
5.7	Panel superior: Periodograma en escala logarítmica de la serie sin huecos (en azul) comparada con el periodograma de la serie interpolada con ARMA (en rojo) y el correspondiente a la serie interpolada linealmente (en verde). En el inserto se muestran estas dos últimas en el rango marcado por una flecha de dos puntas que se corresponde a las frecuencias de 9 a 20 d^{-1} . Panel inferior: Cociente entre las amplitudes obtenidas para la serie interpolada linealmente y las que se obtienen al interpolar los huecos con ARMA. Nótese como las diferencias entre ambos periodogramas se hacen significativas a partir de los 6 d^{-1}	66
5.8	Realización de un modelo analítico compuesto por una señal de altas frecuencias y ruido blanco gaussiano.	67
5.9	Estimación del orden ARMA(p,q) para el modelo que se utiliza en la reconstrucción de los huecos de la serie sintética compuesta por una señal de altas frecuencias y ruido blanco gaussiano.	68

5.10	Comparación entre el método ARMA para el relleno de huecos e interpolación lineal. En el panel superior se muestra un segmento de 0.5 días de la serie generada con el modelo analítico de altas frecuencias. Los segmentos marcados en rojo son los que han sido interpolados en el panel central (interpolación lineal) y en el inferior (interpolación ARMA).	69
5.11	Periodograma en escala logarítmica de la serie interpolada linealmente (en verde) comparada con el periodograma de la serie con huecos (azul).	70
5.12	Periodograma en escala logarítmica de la serie interpolada con ARMA(en rojo) comparada con el periodograma de la serie con huecos (azul). Nótese que en este caso las amplitudes de los frecuencias principales son en torno a un 10% mayores que en la figura anterior.	71
5.13	Panel superior: Periodograma en escala logarítmica de la serie sin huecos (en azul) comparada con el periodograma de la serie interpolada con ARMA(en rojo) y el correspondiente a la serie interpolada linealmente (en verde). Panel inferior: Cociente entre las amplitudes obtenidas para la serie interpolada linealmente y las que se obtienen al interpolar los huecos con ARMA. Nótese como las diferencias entre ambos periodogramas son significativas en todo el rango por debajo de los 400 d^{-1}	72
5.14	Estimación del orden ARMA(p,q) para un modelo de señal con 31 componentes armónicas y ruido blanco gaussiano. Nótese que el modelo óptimo es el ARMA(47,44).	75
5.15	Comparación entre el método ARMA para el relleno de huecos e interpolación lineal para la serie del modelo con 31 frecuencias. En el panel superior se muestra un segmento de 0.4 días y en el panel central e inferior el mismo tras el relleno de huecos de los segmentos marcados en rojo y en cian con el método de interpolación lineal y el método ARMA respectivamente.	76
5.16	Periodograma de la serie sin huecos del modelo de 31 frecuencias (en azul) comparada con el periodograma de la serie con huecos interpolados con ARMA (en rojo) y linealmente (en verde).	77
5.17	Periodicidades en el rango de 0.1 a 60 μHz de la distribución de frecuencias detectadas en los periodograma de la serie sin huecos (en azul) comparada con las periodicidades del periodograma de la serie con huecos interpolados con ARMA(en rojo) y linealmente (en verde). En el inserto se representa la región de bajas frecuencias donde se haya la pequeña separación $\delta\nu$. Los valores reales se calculan como el doble de los hallados en la figura.	81

ÍNDICE DE FIGURAS

5.18	Curva de luz de la estrella de tipo solar HD 49933. Los puntos marcados en rojo se corresponden con los datos que han sido interpolados linealmente en el pipeline de los datos de CoRoT de nivel 2. En el inserto se representa un segmento de un día.	82
5.19	Comparación entre el método ARMA para el relleno de huecos (en rojo) e interpolación lineal (en verde) para la curva de luz de la estrella HD 49933 observada por CoRoT. En los paneles se muestran los huecos más grandes de la serie. Cada panel muestra un segmento distinto de 2 días de duración.	84
5.20	Periodograma en escala logarítmica de la curva de luz de la estrella HD 49933 interpolada linealmente en los huecos (en azul) comparada con el periodograma de la serie con huecos interpolados con ARMA(en rojo). En el inset se muestra el rango de 1100 a 2300 μHz donde se encuentran los modos-p solares. Nótese que las amplitudes de espectro interpolado mediante ARMA son mayores que las del espectro interpolado linealmente.	85
5.21	Curva de luz de la estrella δ Sct HD 174966. Los puntos marcados en rojo se corresponden con los datos que han sido interpolados linealmente por el pipeline que procesa los datos de CoRoT de nivel 2. . . .	86
5.22	Comparación entre el método ARMA para el relleno de huecos (en rojo) e interpolación lineal (en verde) para la curva de luz de la estrella HD 174966 observada por CoRoT.	86
5.23	Periodograma en escala logarítmica de la curva de luz de la estrella HD 174966 interpolada linealmente en los huecos (en azul) comparada con el periodograma de la serie con huecos interpolados con ARMA(en rojo). Nótese como el efecto de la ventana espectral se mantiene con la interpolación lineal.	87
5.24	Periodicidades de HD 174966 en el rango de 0.1 a 100 μHz de la distribución de frecuencias detectadas en los periodograma de la serie interpolada con ARMA (panel superior) comparada con las periodicidades del periodograma de la serie con huecos interpolados linealmente (panel inferior). Las líneas negra, roja y verde se corresponden respectivamente con los subsets de 30, 60 y 114 frecuencias. En el inserto se representa solamente el rango de bajas frecuencias.	88
5.25	60 días de la curva de luz de la estrella Be HD 51193 observada por CoRoT.	89

5.26	Comparación entre el método ARMA para el relleno de huecos (en rojo) e interpolación lineal (en verde) para la curva de luz de la estrella Be HD 51193 observada por CoRoT.	90
5.27	Periodograma en escala logarítmica de la curva de luz de la estrella HD 51193 interpolada linealmente en los huecos (en azul) comparada con el periodograma de la serie con huecos interpolados con ARMA (en rojo). Nótese como el efecto de la ventana espectral se mantiene con la interpolación lineal.	91
6.1	Serie de datos aleatorios generados según una distribución gaussiana de tipo $\mathcal{N}(0, 1)$ normalizada a la unidad. El valor de la serie numérica se representa en ordenadas y en abscisas el índice correspondiente. Nótese como los datos se distribuyen de manera suave en torno a la media. . .	98
6.2	Serie de datos aleatorios generados según una distribución de tipo $\mathcal{B}(0.5, 0.5)$ en el intervalo $[0, 1]$. Nótese como los datos se apilan en los extremos del intervalo.	101
8.1	Segmento de 1000 puntos de la curva de luz de la estrella HD 174936. El panel superior muestra una realización del modelo analítico que ajusta la curva de luz de esta estrella. El panel inferior muestra los datos originales tomados por CoRoT.	113
8.2	Conectividades de la curva de luz de la estrella HD 174936. En el panel superior se muestran las conectividades calculadas con la aproximación ARMA, en el panel intermedio se muestran las conectividades spline, y en el panel inferior los datos originales de CoRoT para su comparación. Nótese las diferentes escalas.	115
8.3	Conectividades para el modelo analítico-1 de la curva de luz de HD 174936. En el panel superior se muestran las conectividades calculadas con la aproximación ARMA, en el panel intermedio se muestran las conectividades spline, y en el panel inferior una realización del modelo analítico para su comparación. Nótese las diferentes escalas.	115
8.4	Conectividades para el modelo analítico-2 de la curva de luz de HD 174936. En rojo se muestran las conectividades calculadas con la aproximación ARMA, en verde se muestran las conectividades spline, y en azul los puntos correspondientes a una realización de este modelo.	116
8.5	Tendencia asintótica del coeficiente SSE con el número de puntos N_p a partir de las conectividades ARMA (en rojo), y las conectividades spline (en azul) para los datos de CoRoT(panel superior) y para el modelo analítico-1 (panel inferior).	117

ÍNDICE DE FIGURAS

8.6	Tendencia asintótica del coeficiente SSE con el número de puntos N_p a partir de las conectividades ARMA (en rojo), y las conectividades spline (en azul) para el modelo analítico-2	118
8.7	Conectividades de la curva de luz de la estrella KIC 006187665. En el panel superior se muestran las conectividades calculadas con la aproximación ARMA, en el panel intermedio se muestran las conectividades spline, y en el panel inferior los datos originales de <i>Kepler</i> para su comparación. Nótese las diferentes escalas.	119
8.8	Tendencia asintótica del coeficiente SSE con el número de puntos N_p a partir de las conectividades ARMA (en rojo), y las conectividades spline (en azul) para los datos de la estrella KIC 006187665.	120
8.9	Conectividades splines de orden-2 (panel inferior) y conectividades spline de orden-1 (panel superior) para la estrella KIC 006187665.	121
8.10	Espectro de las conectividades ARMA (panel superior), conectividades spline (panel intermedio) y de la serie original (panel inferior) de la estrella híbrida KIC 006187665.	122
8.11	Conectividades de la serie de observaciones del Sol. En el panel superior, la aproximación ARMA en negro y la aproximación de splines en gris. En el panel inferior, la serie de datos original de SoHO/GOLF para ser comparada.	123
8.12	Tendencia asintótica del coeficiente SSE con el número de puntos N_p a partir de las conectividades ARMA (en rojo), y las conectividades spline (en azul) para los datos de SoHO/GOLF.	124
8.13	Espectro de Fourier de las conectividades ARMA (panel superior), splines (panel medio), y espectro de Fourier de la serie de datos original (panel inferior) de SoHO/GOLF.	125
8.14	Conectividades de la serie de observaciones de la estrella HD 181072 por el satélite CoRoT. En el panel superior se muestra la aproximación ARMA, en el panel medio la aproximación de splines, y en el panel inferior, la serie de datos original para ser comparada.	126
8.15	Tendencia asintótica del coeficiente SSE con el número de puntos N_p a partir de las conectividades ARMA (en rojo), y las conectividades spline (en azul) para los datos de la estrella HD 181072 observada por CoRoT.	127
8.16	Espectro de Fourier de las conectividades ARMA (panel superior), splines (panel medio), y espectro de Fourier de la serie de datos original (panel inferior) de la estrella constante HD 181072 observada por CoRoT	128

A.1 La ventana Taylor en el dominio del tiempo y su respuesta en frecuencias (azul) comparada con una ventana Hamming (verde) y una Flattop (rojo). Factor de leakage de 0%, atenuación relativa de los lóbulos de -74.8 dB, anchura del pico a -3 dB igual a 0.050781. 142

Índice de tablas

2.1	Clasificación de estrellas pulsantes. Nótese que los intervalos son aproximados.	17
2.2	Técnicas para el análisis de series temporales.	20
3.1	Propiedades de la ventana espectral de una serie con muestreo uniforme respecto a una serie con muestreo no uniforme	40
5.1	Parámetros del modelo usado para el test de bajas frecuencias del método de interpolación ARMA donde A_n son amplitudes, f_n son las frecuencias, y ϕ_n las fases	60
5.2	Parámetros del modelo usado para el test de altas frecuencias del método de interpolación ARMA donde A_n son amplitudes, f_n son las frecuencias, y ϕ_n las fases	65
5.3	Amplitudes estimadas mediante un ajuste de mínimos cuadrados no-lineal a partir de la serie generada con el modelo de frecuencias altas sin huecos A_n^0 , con huecos A_n^{gap} , con interpolación lineal A_n^{lin} , y con interpolación ARMA A_n^{ar}	70
5.4	Fases estimadas mediante un ajuste de mínimos cuadrados no-lineal a partir de la serie generada con el modelo de frecuencias altas sin huecos ϕ_n^0 , con huecos ϕ_n^{gap} , con interpolación lineal ϕ_n^{lin} , y con interpolación ARMA ϕ_n^{ar}	70
5.5	Frecuencias estimadas por medio de un ajuste de mínimos cuadrados no-lineal a partir de la serie generada con el modelo de 31 frecuencias sin huecos f_n^0 , con interpolación ARMA f_n^{ar} , con huecos f_n^{gap} , y con interpolación lineal f_n^{lin} , y las amplitudes y fases correspondientes.	79
5.6	Continuación de la tabla 5.5	80
8.1	Resumen de las estrellas analizadas.	111

- 8.2 Coeficientes de correlación de Pearson (ρ) y las probabilidades de no correlación (P) asociadas a las conectividades ARMA y splines de las series temporales estudiadas en este trabajo. Los subíndices 'a' y 's' se corresponden a las aproximaciones ARMA y spline, respectivamente. . 129

Referencias

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control*, AC-19:716–723. [51](#)
- Anderson, E. (1993). Gap Filling the GONG Data Set. In Brown, T. M., editor, *GONG 1992. Seismic Investigation of the Sun and Stars*, volume 42 of *Astronomical Society of the Pacific Conference Series*, page 445. [46](#)
- Appourchaux, T. (2011). A crash course on data analysis in asteroseismology. *ArXiv e-prints*. [31](#)
- Appourchaux, T., Gizon, L., and Rabello-Soares, M.-C. (1998). The art of fitting p-mode spectra. I. Maximum likelihood estimation. *Astronomy and Astrophysics Supplement*, 132:107–119. [20](#)
- Appourchaux, T., Michel, E., Auvergne, M., et al. (2008). CoRoT sounds the stars: p-mode parameters of Sun-like oscillations on HD 49933. *Astronomy & Astrophysics*, 488:705–714. [42](#), [83](#)
- Auvergne, M. et al. (2009). The CoRoT satellite in flight: description and performance. *Astronomy & Astrophysics*, 506:411–424. [18](#)
- Ballot, J., Gizon, L., Samadi, R., et al. (2011). Accurate p-mode measurements of the G0V metal-rich CoRoT target HD 52265. *Astronomy & Astrophysics*, 530:A97. [42](#)
- Benomar, O., Baudin, F., Campante, T. L., et al. (2009). A fresh look at the seismic spectrum of HD49933: analysis of 180 days of CoRoT photometry. *Astronomy & Astrophysics*, 507:L13–L16. [42](#), [83](#)
- Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day. [108](#)

REFERENCIAS

- Breger, M., Lenz, P., Antoci, V., et al. (2005). Detection of 75+ pulsation frequencies in the δ Scuti star FG Virginis. *Astronomy and Astrophysics*, 435:955–965. [83](#)
- Breger, M., Stich, J., and Garrido, R. (1993). Nonradial Pulsation of the Delta-Scuti Star Bu-Cancri in the Praesepe Cluster. *Astronomy & Astrophysics*, 271:482. [60](#), [142](#)
- Bretthorst, G. L. (2000). Nonuniform Sampling: Aliasing and Bandwidth. In *American Astronomical Society Meeting Abstracts*, volume 32 of *Bulletin of the American Astronomical Society*, page 1438. [37](#)
- Brown, T. M. and Christensen-Dalsgaard, J. (1990). A technique for estimating complicated power spectra from time series with gaps. *ApJ*, 349:667–674. [46](#)
- Burden, R. L. and Faires, J. D. (2011). *Numerical Analysis, 9th ed.* Brooks/Cole. [106](#), [122](#)
- Burg, J. P. (1975). *Maximum Entropy Spectral Analysis*. PhD thesis, University of Stanford. [20](#), [44](#), [49](#)
- Buzasi, D., Catanzarite, J., Laher, R., et al. (2000). The Detection of Multimodal Oscillations on α Ursae Majoris. *The Astrophysical Journal Letters*, 532:L133–L136. [18](#)
- Carbonell, M., Oliver, R., and Ballester, J. L. (1992). Power spectra of gapped time series: a comparison of several methods. *Astronomy and Astrophysics*, 264:350–360. [45](#)
- Carleson, L. (1966). On convergence and growth of partial sums of fourier series. *Acta Mathematica*, 116(1):135–157. [102](#)
- Chapellier, E., Rodríguez, E., Auvergne, M., et al. (2011). The γ doradus corot target hd 49434. ii. frequency analysis of the corot data. *Astronomy and Astrophysics*, 525:A23. [15](#), [95](#), [122](#), [149](#)
- Christensen-Dalsgaard, J. (2003). *Lecture notes on Stellar Oscillations, 5th edn.* Institut for Fysik og Astronomi, Aarhus Universitet. [13](#), [149](#)
- Christensen-Dalsgaard, J. (2004). An Overview of Helio- and Asteroseismology. volume 559 of *ESA Special Publication*, page 1. [11](#)
- Cincotta, P. M., Mendez, M., and Nunez, J. A. (1995). Astronomical Time Series Analysis. I. A Search for Periodicity Using Information Entropy. *The Astrophysical Journal*, 449:231. [20](#)

- Cowling, T. G. (1941). The non-radial oscillations of polytropic stars. *MNRAS*, 101:367. [12](#)
- De Ridder, J., Barban, C., Baudin, F., et al. (2009). Non-radial oscillation modes with long lifetimes in giant stars. *Nature*, 459:398–400. [18](#)
- Deeming, T. J. (1975). Fourier analysis with unequally spaced data. *ApSS*, 36:137–158. [20](#), [32](#), [35](#)
- Degroote, P., Briquet, M., Catala, M., et al. (2009). Evidence for nonlinear resonant mode coupling in the β cephei star hd 180642 (v1449 aquilae) from corot photometry. *Astronomy and Astrophysics*, 506:111–123. [100](#), [111](#)
- Deheuvels, S., Bruntt, H., Michel, E., et al. (2010). Seismic and spectroscopic characterization of the solar-like pulsating CoRoT target HD 49385. *Astronomy and Astrophysics*, 515:A87. [42](#)
- Dziembowski, W. (1982). Nonlinear mode coupling in oscillating stars. i - second order theory of the coherent mode coupling. *Acta Astronómica*, 32:147. [100](#)
- Eddington, A. S. (1918). Stars, Gaseous, On the pulsations of a gaseous star. *MNRAS*, 79:2–22. [11](#)
- Elad, M., Starck, J.-L., Querre, P., and Donoho, D. (2005). Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Applied and Computational Harmonic Analysis*, 19(3):340 – 358. [43](#)
- Esquef, P. A. A. and Biscainho, L. W. P. (2006). An efficient model-based multirate method for reconstruction of audio signals across long gaps. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(4):1391–1400. [49](#)
- Esquef, P. A. A., Välimäki, V., Roth, K., and Kauppinen, I. (2003). Interpolation of long gaps in audio signals using the warped burg’s method. In *Proceedings of the 6th Int. Conference on Digital Audio Effects (DAFx-03)*, pages 1–6. [49](#)
- Fahlman, G. G. and Ulrych, T. J. (1982). A new method for estimating the power spectrum of gapped data. *Mon. Not. R. Astr. Soc.*, 199:53–65. [44](#), [50](#)
- Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proc. Roy. Soc. Ser. A*, 125:54–59. [99](#)
- Fossat, E., Kholikov, S., Gelly, B., et al. (1999). Full disk helioseismology: repetitive music and the question of gap filling. *Astronomy & Astrophysics*, 343:608–614. [46](#)

REFERENCIAS

- García, R. A., Hekker, S., Stello, D., et al. (2011). Preparation of Kepler light curves for asteroseismic analyses. *MNRAS*, 414:L6–L10. [33](#)
- García, R. A., Régulo, C., Samadi, R., et al. (2009). Solar-like oscillations with low amplitude in the CoRoT target HD 181906. *Astronomy & Astrophysics*, 506:41–50. [42](#)
- García Hernández, A., Moya, A., Michel, E., et al. (2009). Asteroseismic analysis of the CoRoT δ Scuti star HD 174936. *Astronomy and Astrophysics*, 506:79–83. [95](#), [111](#), [112](#)
- García Hernández, A., Moya, A., Michel, E., et al. (2013). An in-depth study of HD 174966 with CoRoT photometry and HARPS spectroscopy. Large separation as a new observable for δ Sct stars. *ArXiv e-prints*. [42](#), [83](#), [87](#)
- García, R. A., Turck-Chieze, S., Boumier, P., et al. (2005). Global solar doppler velocity determination with the golf/soho instrument. *Astronomy & Astrophysics*, 442:385–395. [111](#), [123](#)
- García Hernández, A. (2011). *Cuasi periodicidades en los periodogramas de estrellas δ Scuti: un nuevo observable. Dos ejemplos de CoRoT: HD 174936 y HD 174966*. PhD thesis, Universidad de Granada. [74](#), [78](#), [87](#)
- Garrido, R. (2000). Photometric Modal Discrimination in δ Scuti and γ Doradus Stars. In Breger, M. and Montgomery, M., editors, *Delta Scuti and Related Stars*, volume 210 of *Astronomical Society of the Pacific Conference Series*, page 67. [35](#)
- Garrido, R. and Poretti, E. (2004). The study of δ Scuti stars in the transition era from ground-based to space photometry. In *IAU Colloq. 193: Variable Stars in the Local Group*, volume 310 of *Astronomical Society of the Pacific Conference Series*, page 560. [83](#)
- Garrido, R. and Rodríguez, E. (1996). Microvariability in high-amplitude delta scuti radially pulsating stars. *MNRAS*, 281:696. [100](#)
- Gilliland, R. L., Brown, T. M., Christensen-Dalsgaard, J., et al. (2010). Kepler Asteroseismology Program: Introduction and First Results. *Pub. Astron. Soc. Pac.*, 122:131–143. [18](#), [118](#)
- Gizon, L. and Solanki, S. K. (2003). Determining the Inclination of the Rotation Axis of a Sun-like Star. *The Astrophysical Journal*, 589:1009–1019. [78](#)

- Grigahcène, A., Dupret, M.-A., Gabriel, M., et al. (2005). Convection-pulsation coupling. I. A mixing-length perturbative theory. *Astronomy & Astrophysics*, 434:1055–1062. [16](#)
- Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23(2):347–356. [97](#)
- Gutiérrez-Soto, J., Fabregat, J., Suso, J., et al. (2007). A photometric study of Be stars located in the seismology fields of COROT. *Astronomy and Astrophysics*, 476:927–933. [89](#)
- Gutiérrez-Soto, J., Floquet, M., Samadi, R., et al. (2009). Low-amplitude variations detected by CoRoT in the B8IIIe star HD 175869. *Astronomy and Astrophysics*, 506:133–141. [42](#)
- Harvey, J. (1985). High-resolution helioseismology. *Proc. ESA Workshop on Future Missions in Solar, Heliospheric & Space Plasma Physics*, pages 199–208. [113](#)
- Högbom, J. A. (1974). Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astronomy & Astrophysics Suppl.*, 15:417. [20](#), [45](#)
- Holschneider, M., Kronland-Martinet, R., Morlet, J., and Tchamitchian, P. (1990). A real-time algorithm for signal analysis with the help of the wavelet transform. In Combes, J.-M., Grossmann, A., and Tchamitchian, P., editors, *Wavelets*, inverse problems and theoretical imaging, pages 286–297. Springer Berlin Heidelberg. [42](#)
- Jech, T. J. (1997). *Set Theory, 2nd ed.* Springer-Verlag. [103](#)
- Kallinger, T. and Matthews, J. M. (2010). Evidence for granulation in early a-type stars. *The Astrophysical Journal*, 711:L35–L39. [112](#), [113](#)
- Kaplan, w. (1992). *Advanced Calculus, 4th ed.* Addison-Wesley. [102](#)
- Korzennik, S. G., Rabello-Soares, M. C., Schou, J., and Larson, T. P. (2013). Accurate Characterization of High-degree Modes Using MDI Observations. *The Astrophysical Journal*, 772:87. [45](#)
- Kovács, G., Zucker, S., and Mazeh, T. (2002). A box-fitting algorithm in the search for periodic transits. *Astronomy and Astrophysics*, 391:369–377. [20](#)
- Kovács, G. (1983). On the accuracy of frequency determination by an autoregressive spectral estimator. *Solar Physics*, 82:123–128. [51](#)
- Ledoux, P. (1951). The Nonradial Oscillations of Gaseous Stars and the Problem of Beta Canis Majoris. *The Astrophysical Journal*, 114:373. [12](#)

REFERENCIAS

- Lenz, P. and Breger, M. (2005). Period04 User Guide. *Communications in Asteroseismology*, 146:53–136. [21](#), [57](#), [85](#)
- Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39:447–462. [20](#)
- Mantegazza, L., Poretti, E., Michel, E., et al. (2012). Pulsation spectrum of δ Scuti stars: the binary HD 50870 as seen with CoRoT and HARPS. *Astronomy & Astrophysics*, 542:A24. [42](#)
- Mathur, S., A., G. R., Catalá, C., et al. (2010). The solar-like corot targe hd 170987: spectroscopic and seismic observations. *Astronomy & Astrophysics*, 518:A53. [43](#)
- Mathur, S., Bruntt, H., Catala, C., et al. (2013). Study of HD 169392A observed by CoRoT and HARPS. *Astronomy & Astrophysics*, 549:A12. [43](#)
- Mathur, S., Garcia, R. A., Regulo, C., et al. (2010). An automatic pipeline analysing solar-like oscillating targets tested on CoRoT and simulated data. *ArXiv e-prints*. [43](#)
- Matthews, J. M., Kuschnig, R., Guenther, D. B., et al. (2004). No stellar p-mode oscillations in space-based photometry of Procyon. *Nature*, 430:51–53. [18](#)
- Montgomery, M. and O’Donoghue, D. (1999). A derivation of the errors for least squares fitting to time series data. *Delta Scuti Star Newsletter*, 13:28. [142](#)
- Mosser, B., Michel, E., Appourchaux, T., et al. (2009). The CoRoT target HD 175726: an active star with weak solar-like oscillations. *Astronomy & Astrophysics*, 506:33–40. [42](#)
- Moya, A. and Garrido, R. (2008). Granada oscillation code (graco). *Astrophys. Space Sci.*, 316:129–133. [16](#)
- Neiner, C., Gutiérrez-Soto, J., Baudin, F., et al. (2009). The pulsations of the B5IVe star HD 181231 observed with CoRoT and ground-based spectroscopy. *Astronomy and Astrophysics*, 506:143–151. [42](#)
- Paparo, M. and Kovacs, G. (1984). FM COM - Has this Delta Scuti-star variable frequency spectra? *Astrophysics & Space Science*, 105:357–367. [45](#)
- Parks, T. W. and Burrus, C. S. (1987). *Digital Filter Design*. John Wiley & Sons. [53](#)
- Pascual-Granado, J., Garrido, R., Gutiérrez-Soto, J., and Martín-Ruiz, S. (2012). Towards a more general method for filling gaps in time series. In *New Horizons in Time Domain Astronomy*, volume 285 of *IAU Symposium*, pages 392–393. [30](#), [50](#)

- Pascual-Granado, J., Garrido, R., and Suárez, J. C. (2014). On the necessity of a new interpretation of the stellar light curves. In *Precision Asteroseismology*, pages 85–88. [95](#), [111](#)
- Pascual-Granado, J., Grigahcène, A., Díaz-Fraile, D., et al. (2013). On the origin of the dense frequency spectra of space observed intermediate mass pulsating stars. In *Stellar Pulsations: Impact of New Instrumentation and New Insights*, volume 31, page 33. [57](#)
- Petersen, J. O. (1973). Masses of double mode cepheid variables determined by analysis of period ratios. *Astronomy & Astrophysics*, 27:89–93. [73](#)
- Pires, S., Starck, J.-L., Amara, A., et al. (2009). FAst STatistics for weak Lensing (FASTLens): fast method for weak lensing statistics and map making. *MNRAS*, 395:1265–1279. [43](#)
- Poretti, E., Michel, E., Garrido, R., et al. (2009). Hd 50844: a new look at δ scuti stars from corot space photometry. *Astronomy and Astrophysics*, 506:85–93. [95](#), [122](#)
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C (2Nd Ed.): The Art of Scientific Computing*. Cambridge University Press. [21](#)
- Priestley, M. B. (1981). *Spectral Analysis and Time Series Volume 1*. Academic Press. [99](#)
- Priestley, M. B. (1988). *Non-linear and non-stationary time series analysis*. Academic Press. [101](#), [129](#)
- Rauer, H., Queloz, D., Csizmadia, S., et al. (2009). Transiting exoplanets from the CoRoT space mission. VII. The “hot-Jupiter”-type planet CoRoT-5b. *Astronomy and Astrophysics*, 506:281–286. [42](#)
- Reegen, P. (2007). SigSpec. I. Frequency- and phase-resolved significance in Fourier space. *Astronomy & Astrophysics*, 467:1353–1371. [20](#), [57](#), [85](#)
- Régulo, C., Vázquez Ramió, H., and Roca Cortés, T. (2005). An observational approach to convection in main sequence stars. *Astronomy and Astrophysics*, 443:1013–1019. [20](#)
- Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, volume 293 of *A series of Comprehensive Studies in Mathematics*. Springer. [105](#)

REFERENCIAS

- Roques, S., Schwarzenberg-Czerny, A., and Dolez, N. (2000). Parametric Spectral Analysis Applied to Gapped Time-Series of Variable Stars. *Baltic Astronomy*, 9:463–477. [50](#)
- Roth, M. and Zhugzhda, Y. D. (2010). Gapfilling interrupted helioseismic data with the EM algorithm. *Astronomy Letters*, 36:64–73. [47](#)
- Roxburgh, I. W. and Vorontsov, S. V. (1994). The Asymptotic Theory of Stellar Acoustic Oscillations - a Fourth-Order Approximation for Low-Degree Modes. *MNRAS*, 268:143. [74](#)
- Royden, H. L. (1988). *Real Analysis*. Prentice-Hall. [107](#)
- Saio, H., Baker, N. H., and Gautschy, A. (1998). On the origin of strange modes. *MNRAS*, 294:622. [100](#)
- Salakhutdinova, I. I. (1998). A Fractal Structure of the Time Series of Global Indices of Solar Activity. *Solar Physics*, 181:221–235. [45](#)
- Samadi, R., Fialho, F., Costa, J. E. S., et al. (2007). The corot book: Chap. v.5/ extraction of the photometric information : corrections. *ArXiv Astrophysics e-prints*. [34](#), [42](#), [114](#)
- Sato, K. H., Garcia, R. A., Pires, S., et al. (2010). Inpainting: A powerful interpolation technique for helio- and asteroseismic data. *ArXiv e-prints*. [44](#)
- Scargle, J. D. (1981). Studies in astronomical time series analysis. I - Modeling random processes in the time domain. *Astrophysical Journal Supplement Series*, 45:1–71. [51](#), [53](#)
- Scargle, J. D. (1982). Studies in astronomical time series analysis. ii. statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853. [20](#), [36](#), [95](#)
- Scargle, J. D. (1990). Studies in astronomical time series analysis. iv - modeling chaotic and random processes with linear filters. *The Astrophysical Journal*, 359:359. [108](#)
- Schwarzenberg-Czerny, A. (1987). On the X-Ray Variability of Tt-Arietis. *Acta Astronomica*, 37:213. [45](#)
- Schwarzenberg-Czerny, A. (1996). Fast and Statistically Optimal Period Search in Uneven Sampled Observations. *Astrophysical Journal Letters*, 460:L107. [20](#)

- Schwarzenberg-Czerny, A. (2003). An astronomer's guide to period searching. *ASP Conference series*, 292:383–390. [107](#)
- Shannon, C. E. (1949). Communication in the presence of noise. *Proc. IEEE*, 37:10–21. [31](#), [104](#)
- Shibahashi, H. (1979). Modal analysis of stellar nonradial oscillations by an asymptotic method. *Publ. Astron. Soc. Japan*, 31:87. [14](#)
- Spiegel, M. R. (1992). *Theory and Problems of Probability and Statistics*. McGraw-Hill. [97](#)
- Starck, J.-L., Fadili, M. J., and Rassat, A. (2013). Low- ℓ CMB analysis and inpainting. *ApJ*, 550:A15. [43](#)
- Steiglitz, K. and McBride, L. E. (1965). A technique for the identification of linear systems. *IEEE Trans. Automatic Control*, AC-10:461–464. [53](#)
- Stellingwerf, R. F. (1978). Period determination using phase dispersion minimization. *The Astrophysical Journal*, 224:953–960. [20](#)
- Stoica, P., Larsson, E. G., and Li, J. (2000). Adaptive Filter-bank Approach to Restoration and Spectral Analysis of Gapped Data. *The Astronomical Journal*, 120:2163–2173. [48](#)
- Suárez, J. C., García Hernández, A., Moya, A., et al. (2014). Measuring mean densities of δ Scuti stars with asteroseismology. Theoretical properties of large separations using TOUCAN. *Astronomy & Astrophysics*, 563:A7. [83](#)
- Suárez, J. C., Goupil, M. J., and Morel, P. (2006). Effects of moderately fast shellular rotation on adiabatic oscillations. *Astronomy & Astrophysics*, 449:673–685. [16](#)
- Tassoul, M. (1980). Asymptotic approximations for stellar nonradial pulsations. *ApJS*, 43:469. [14](#), [71](#)
- Toner, C. G., Jefferies, S. M., and Toutain, T. (1999). Increasing the Visibility of Solar Oscillations. *ApJL*, 518:L127–L130. [46](#)
- Ulrych, T. J. and Clayton, R. W. (1976). Time series modelling and maximum entropy. *Physics of the Earth and Planetary Interiors*, 12(2–3):188 – 200. [49](#), [51](#)
- Ulrych, T. J. and Ooe, M. (1979). Autoregressive and Mixed Autoregressive-Moving Average Models and Spectra (With 13 Figures). In Haykin, S., editor, *Nonlinear Methods of Spectral Analysis*, page 73. [49](#), [51](#), [61](#)

REFERENCIAS

- Uytterhoeven, K., Mathias, P., Baglin, A., et al. (2011). The variability of the CoRoT target HD171834: gamma Dor pulsations and/or activity? *ArXiv e-prints*. [43](#)
- Uytterhoeven, K., Moya, A., Guzik, J. A., et al. (2011). Characterization of the variability amongst a- and f-type stars. *Astronomy & Astrophysics*, 534:70. [57](#), [111](#), [118](#)
- van Dijk, G. (2009). *Introduction to harmonic analysis and generalized Gelfand pairs*. Walter de Gruyter. [102](#)
- Vaniček, P. (1971). Further Development and Properties of the Spectral Analysis by Least-Squares. *Astrophysics and Space Science*, 12:10–33. [21](#)
- Ventkus, V. Y. (1982). Analyticity of gaussian measures. *Theory of Probability and its Applications*, 27(1):155–161. [101](#)
- Wiener, N. (1923). Differential space. *J. Math. and Phys.*, 2:131–174. [106](#), [116](#)
- Wold, H. (1938). *A study in the analysis of stationary time series 2nd ed.* Almqvist and Wiksell. [48](#), [104](#)
- Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer’s sunspot numbers. *Royal Society of London Philosophical Transactions Series A*, 226:267–298. [47](#)
- Zechmeister, M. and Kürster, M. (2009). The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. *Astronomy and Astrophysics*, 496:577–584. [20](#)
- Zhevakin, S. (1953). K teorii cefeid. i. *Astron. Zh.*, 30:161–179. [11](#)
- Zwintz, K., Kallinger, T., Guenther, D. B., et al. (2011). Pulsational Analysis of V 588 Mon and V 589 Mon Observed with the MOST and CoRoT Satellites. *The Astrophysical Journal*, 729:20. [42](#)