

# ANÁLISIS ESPECTRAL DE OLAS MARINAS: MODELOS UNIVARIADOS

\* Nestor Escudero Mora      \*\* Joaquín Ortega Sánchez  
\*\*\* Jhonny Escalona

Recibido: 03/03/2015    Aprobado: 28/01/2016

## Resumen

El análisis espectral, es el proceso técnico de la descomposición de una señal compleja en partes más simples. Muchos procesos físicos se describen mejor como una suma de muchas componentes de frecuencias individuales, alternativamente una señal puede dividirse en segmentos cortos y el análisis del espectro puede aplicarse a estos segmentos individuales. Desde el punto de vista del estudio de olas marinas, el espectro juega un rol fundamental y es interpretado como la energía presente en la serie de tiempo durante el período. Este estudio se basa en el análisis del espectro asociado a datos del mar del Norte en el año de 1999 en la plataforma North Alwyn almacenados en la Universidad Heriot-Watt de Edimburgo, separados en períodos de 20 minutos, el total a estudiar son 244 períodos, para ello, se determinó que existe información que no aporta al problema, representa ruido y es el 3% de la energía que no fue tomada para el estudio. Se dividió el espectro en 12 subintervalos y se ajustaron modelos autorregresivos (AR) bajo el esquema de Box-Jenskin and Reinsel, estudiados estos hasta encontrar la mejor aproximación posible para cada subintervalo mediante las propiedades estadísticas obtenidas por cada modelo, junto con los pronósticos respectivos.

**Palabras clave:** Espectro, análisis espectral, dominio de frecuencias, modelos autorregresivos.

---

\* *Universidad Politécnica Arístides Bastidas, Yaracuy, Venezuela, MSc. en Ciencias, nestj2803@gmail.com*

\*\* *Centro de Investigaciones Matemáticas, Guanajuato, México, PhD en Matemáticas, jortega@cimat.mx*

\*\*\* *Universidad Centroccidental Lisandro Alvarado, Departamento de Investigación de Operaciones y Estadísticas, Barquisimeto, Venezuela. MSc. en Modelos Aleatorios, jhonnyscalona@ucla.edu.ve*

## SPECTRAL ANALYSIS OF WAVE SEA: UNIVARIATE MODELS

### Abstract

Spectral analysis is the technical process of a complex signal decomposition into simpler parts. Many physical processes are better described as a sum of many individual frequency components; alternatively a signal can be divided into short segments and the spectrum analysis can be applied to these individual segments. From the point of view of the study of Waves Sea, the spectrum plays a vital role and it is interpreted as the energy in the time series during the period. This study is based on analysis of the spectrum associated to data from the North Sea in the year of 1999, taken in North Alwyn platform, stored in the Heriot-Watt University in Edinburgh, separated in periods of 20 minutes. The total of the studied periods is 244. Thus, it was determined that there is information that does not add information to the problem, and it represents noise, which is the 3 % of energy and it was not taken for the study. The spectrum was divided into 12 subintervals. Autoregressive models were fitted under the scheme Box-Jenskin and Reinsel. The model was studied until the best possible approximation for each subinterval was found by the statistical properties obtained by each model, together with the respective forecasts.

**Keywords:** Spectrum, spectral analysis, frequency domain, autoregressive, models.

## INTRODUCCIÓN

Se realizó un análisis espectral de 244 períodos de una tormenta en el Mar del Norte ocurrido en 1999. Estos datos fueron registrados a una frecuencia de  $5Hz$ , los espectros obtenidos a partir de los datos de la tormenta pueden apreciarse en la Figura 1, donde se pueden observar importantes fluctuaciones por cada periodo. El área total bajo la curva representa la energía en cada uno de los intervalos de tiempo, la gráfica siguiente está hecha en base al espectro en función de la frecuencia, de donde luego se obtendrá importante información para su posterior estudio. El espectro proporciona información diferente de las propiedades estadísticas de la serie de tiempo, ya que da la distribución de la varianza de la serie con la frecuencia. Como muestran Jenkins y Watts (1968) el espectro puede ser graficado en función de la frecuencia en el rango comprendido entre 0 y  $1/2$  (Lindgren, 2006;

Ortega, 2001; Ochi, 2005).

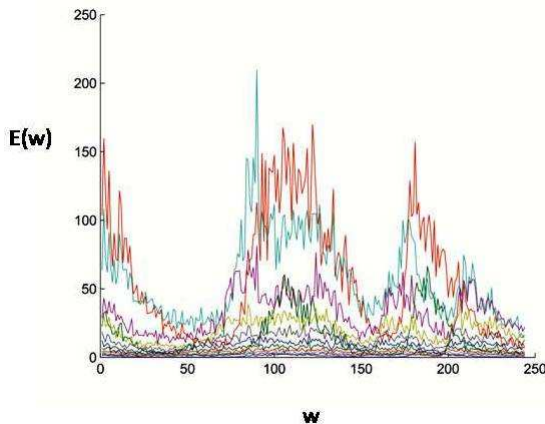


Figura 1: Espectros de los 244 periodos en estudio.

La técnica utilizada para ajustar modelos estocásticos para la evolución espectral fue la siguiente: los registros de altura del mar están divididos en intervalos de 20 minutos y para cada uno de ellos se calculó la densidad espectral usando el software WAFO. Inicialmente, se determinó el *rango útil* de frecuencias común a todos los espectros y para esto se seleccionaron empíricamente valores a los extremos del rango, de modo que no se observan características importantes fuera de este rango y que la energía total presente fuera del rango fuese pequeña. El rango obtenido, luego de este procedimiento, se dividió en intervalos de igual tamaño y se calculó la energía presente en cada intervalo por cada uno de los espectros, de modo que se tiene una sucesión de valores de la energía presentes en cada banda de frecuencias a lo largo de la tormenta.

A cada una de estas series, se le ajustó un modelo AR cuyo orden se determinó usando las herramientas habituales como las gráficas de la ACF y PACF, el  $p$ -valor de los coeficientes y el criterio de información de Akaike (Shumway y Stoffer, 2006). Para verificar el ajuste, se hizo un estudio de los residuales de los modelos obtenidos, usando

la gráfica de cuantiles para comparar el ajuste con una distribución normal. En caso afirmativo, los residuales deben colocarse aproximadamente en una línea recta. Para algunos modelos se encontró que, este no era el caso y se recurrió a la transformación de Box-Cox sobre los datos con el fin de lograr que la hipótesis de Gaussianidad se cumpla, al menos aproximadamente.

## DESARROLLO

### MODELOS ASOCIADOS AL ESTUDIO DE LA ENERGÍA

Toda estimación incluye un error o ruido, y en caso de las densidades espectrales lo que se observa en muy bajas o muy altas frecuencias es esencialmente ruido. El objetivo es reducir el rango de frecuencias dejando fuera aquellas frecuencias en las que se observa principalmente ruido. Luego de examinar las gráficas y de calcular el porcentaje de la energía que queda fuera del intervalo, se decidió tomar el intervalo  $[0,25; 1,45]$ , dentro del cual se encuentra aproximadamente el 97% de la energía total para todos los períodos.

Nos proponemos discretizar el intervalo  $[0,25; 1,45]$  en subintervalos apropiados, de tal manera que estos capturen de la mejor manera posible la variabilidad de la energía a lo largo de la tormenta, pero teniendo en cuenta que un número elevado de intervalos va a producir dificultades en la modelación y su posible interpretación. Luego de observaciones gráficas, se decidió dividir el rango en 12 subintervalos. La gráfica a continuación (Figura 2) muestra la partición para la densidad espectral, la cual tendrá 12 series de tiempo que se obtienen a partir de la división del rango de frecuencias.

Por cada partición, se hizo el estudio del modelo AR apropiado que corresponde a cada uno de estos subintervalos de frecuencia. Esto último se realizó estudiando su función de autocovarianza (ACF) y su función de autocovarianza parcial (PACF), parte importante en este punto, también se estudió el comportamiento de los residuales del modelo respecto a la gráfica de cuantiles para verificar el ajuste con una distribución normal. Otro análisis determinante es la esco-

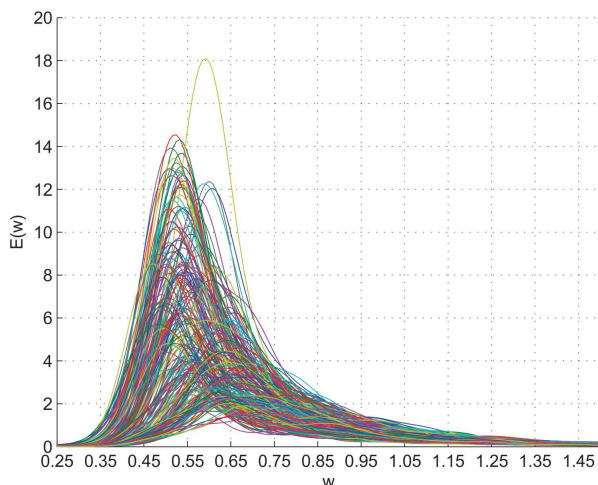


Figura 2: Evolución de la energía para cada uno de los 12 subintervalos.

gencia y bondad de los modelos en estudio, el análisis de los  $p$ -valores mediante el *estadístico de Ljung-Box* (ver por ejemplo Shumway y Stoffer (2006) y Brockwell y Davis (2002)) de acuerdo al cual podemos decidir si el ajuste del modelo resulta apropiado.

### Sobre la construcción de los modelos

En esta sección, se estudian las características que poseen la ACF (Función de Autocorrelación) y la PACF (Función de Autocorrelación Parcial), y a partir de allí se obtiene la respuesta sobre el tipo de modelos y orden de los modelos que se usaron para el presente estudio. A continuación, se presenta un análisis sobre los elementos que se han mencionado anteriormente para la escogencia de los modelos adecuados por cada partición. Por ejemplo, la Figura 3 muestra, para el primer subintervalo, las PACF y ACF para la serie de tiempo correspondiente, donde se evidencia un decaimiento de la ACF hasta el número 18. Por otro lado, la ACP muestra tener los primeros 4

valores fuera del intervalo de confianza, por lo que estas características sugieren un modelo autorregresivo de orden 4. Evidencias de este tipo se analizaron para los demás subintervalos y se obtuvieron órdenes entre 4 y 6, pero este análisis no es suficiente, ya que no hemos considerado todos los componentes estadísticos que posee el modelo (Cowpertwait y Metcalfe, 2008). Haciendo un análisis más exhaustivo y experimentando con los órdenes para el modelo AR correspondiente a cada uno de los subintervalos, se encontró que el orden para cada uno de estos modelos puede reducirse, viendo los coeficientes del modelo, siendo que algunos pueden ser despreciados por su proximidad al cero, y además observando el menor valor para el AIC (Akaike Information Criteria) y el estudio de los  $p$ -valores de cada modelo encontrado (Shumway y Stoffer, 2006).

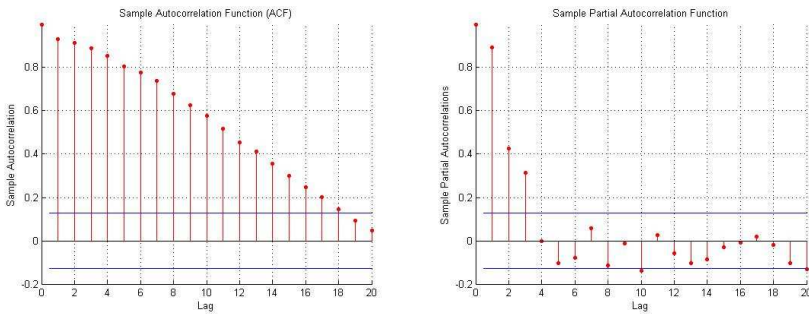


Figura 3: Información sobre el modelo asociado al primer subintervalo. Izquierda: Función de autocorrelación, derecha: Función de autocorrelación parcial

## CARACTERÍSTICAS DE LOS MODELOS

A continuación, se da la justificación de la elección del orden y tipo de los modelos, además se muestra justificación de la escogencia de este orden mostrando la forma estadística de algunas de los componentes del modelo. Por ejemplo, el valor AIC asociado al modelo, las características de sus residuales, su contraste con los cuantiles normales, aunado a esto, las funciones de autocorrelación y autocorrelación parcial.

relación parcial además del estadístico Ljung Box. En este punto, se evidencian algunas particularidades en los residuos, como es el cambio brusco de su comportamiento, esto dado que hay cambios muy marcados positiva y negativamente en los residuales y casi de manera instantánea los siguientes residuos se aproximan a cero. Este fenómeno se conoce como *heterocedasticidad*.

Este tipo de fenómenos, aporta a la no normalidad de los datos porque hay indicios de colas pesadas, lo que impide que los residuos se ajusten a la línea recta deseada, más bien, lo que se vuelve es una función cóncava hacia arriba por un lado y luego cambia a cóncava hacia abajo por otro lado o viceversa. Aunado a esto, la evidencia de que algunos valores de la autocorrelación muestral fallan dentro del intervalo de confianza que considera el estadístico de Ljung-Box, en este punto ya el modelo se considera poco viable y es necesario considerar otras alternativas para detectar la bondad del modelo y no rechazarlo.

## COMPARACIÓN DE LOS MODELOS OBTENIDOS Y SELECCIÓN DE LOS MODELOS CON MEJOR AJUSTE

A continuación, se reflejan gráficamente las componentes de los modelos más adecuados que se han estimado, haciendo cambios ligeros en el orden del modelo AR. Como consecuencia se observan notablemente como los  $p$ -valores para uno de los modelos entran dentro de la banda de confianza para el estadístico de Ljung Box, además los residuales reflejan aleatoriedad, estos son los criterios entre otros, que se han considerado para la elección de los modelos a estudiar. Como una muestra comparamos el modelo AR de orden 2 y de orden 1 para el primer subintervalo, de tal forma que se puedan apreciar las diferencias planteadas sobre el comportamiento de sus componentes estadísticas. Explícitamente tenemos que los modelos son

$$AR(2): x_t = 0,006 + 0,71x_{t-1} + 0,25x_{t-2} + \omega_t$$

$$AR(1): x_t = 0,006 + 0,94x_{t-1} + \omega_t$$

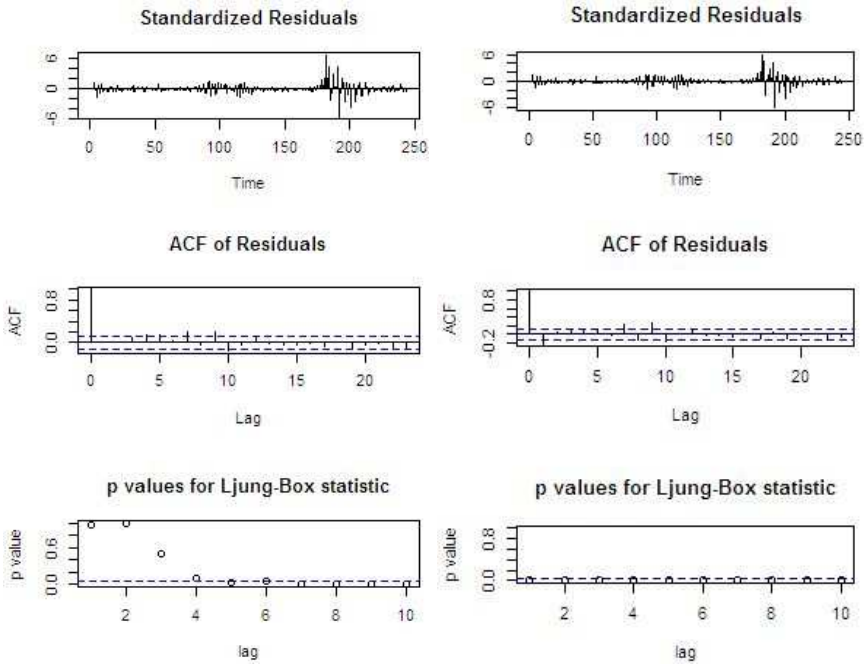


Figura 4: Diagnóstico para el primer subintervalo. Izquierda:  $ARMA(2,0)$ ,  $AIC = -2308,13$ . Derecha:  $ARMA(1,0)$ ,  $AIC = -2294,97$

En la Figura 4, se pueden apreciar los modelos más apropiados que hemos hallado para el primer subintervalo, en este punto se debe decidir sobre cuál de estos modelos es el que mejor ajusta los datos. Para ello notamos que en la gráfica de la izquierda se observa que los residuos normalizados muestran aleatoriedad en su comportamiento, esto da indicios de que no existe una tendencia específica en el comportamiento de los datos, pero al ver el comportamiento de los  $p$ -valores, se observa que algunos de los valores permanecen fuera del intervalo de confianza considerado apropiado para estos, caso contrario ocurre con la figura de la derecha, en la cual los residuos normalizados muestran aún más indicios de aleatoriedad y la prueba para los  $p$ -valores indica que estos todos permanecen dentro del intervalo de confianza para el estadístico que se está usando, por lo que el



modelo que se considerará para el estudio de este subintervalo es un modelo  $AR(1)$  (Cowpertwait y Metcalfe, 2008; Shumway y Stoffer, 2006).

## CARACTERÍSTICAS DE LOS RESIDUOS Y GRÁFICA DE CUANTILES PARA LOS MODELOS

Los componentes importantes en el análisis de datos para conocer y decidir sobre modelos ARMA en general, son parte del estudio exhaustivo de los residuales de estos modelos, ya que dan un panorama claro sobre el comportamiento normal o no de estos últimos, clave en el presente estudio, ya que en definitiva estas gráficas hablan sobre el buen ajuste de los modelos que se están considerando. En este sentido, es conveniente mencionar que en algunas situaciones se presentan particularidades en el comportamiento de los residuales y una de estas, por lo menos la más común, es que se evidencia *heterocedasticidad*. En presencia de este fenómeno la varianza cambia de manera brusca y las oscilaciones se hacen pronunciadas, por lo que al observar la gráfica de cuantiles la distribución de estos es Gaussiana, es decir, los cuantiles deben aproximarse a una línea recta; en caso contrario, si se observa algún comportamiento cóncavo o convexo se hace uso de la *transformación de Box-Cox*.

Hicimos el análisis para el tercer subintervalo de forma precisa y amplia, para luego mostrar a los que sí necesitan de la transformación, los cuales son el primero, segundo, quinto, sexto, séptimo y octavo subintervalo en el orden que están en la partición. En este caso mostraremos el análisis para el primer subintervalo ampliando su estudio para dejar un panorama claro de lo que se pretende hacer con los restantes. De esta manera, observar y entender la técnica empleada para el análisis. Mostramos también las gráficas de cuantiles, quienes nos proporcionarán la información necesaria para aplicar o no la transformación de Box-Cox. La transformación sirve para aproximar la línea recta para obtener la normalidad de los residuos del modelo en el caso que el comportamiento sea el antes descrito, cóncavo o convexo. Se muestran también el comportamiento de los residuales del modelo con la transformación de Box-Cox y sin ella,

de manera tal que se pueda observar las diferencias que existen en el comportamiento de los residuos luego de aplicar la transformación.

### RESIDUALES DEL MODELO ASOCIADO AL TERCER SUBINTERVALO QUE APROXIMAN A LA DISTRIBUCIÓN NORMAL

En esta parte, mostramos aquellas gráficas de cuantiles de los modelos asociados a los subintervalos que no requieren la aplicación de la transformación Box Cox, esto quiere decir como ya se señaló, que el parámetro  $\lambda$  que se requiere para aplicar dicha transformación es aproximadamente 1. Entonces los modelos asociados a cada uno de estos subintervalos proporcionan un buen ajuste y se consideran apropiados para el estudio en cuestión. Particularmente, se muestra el análisis completo para el tercer subintervalo, que como ya se dijo posee estas características. A continuación, la expresión para el modelo  $AR(1)$  asociado al tercer subintervalo se muestra específicamente y seguidamente, la ya mencionada gráfica de cuantiles.

$$x_t = 0,9194 + 0,2236x_{t-1} + \omega_t$$

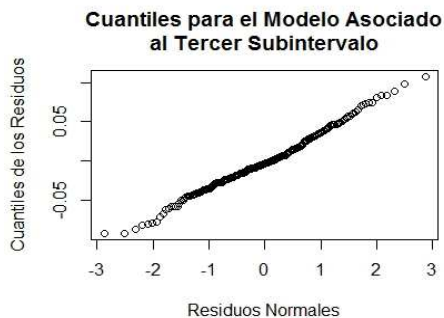


Figura 5: Gráficas de Cuantiles para el tercer subintervalo.

En la Figura 5 se presentan los residuos frente a los cuantiles de la distribución. Si los residuos tienen distribución Normal deben

quedar todos los puntos aproximadamente sobre una recta, este es el caso *ideal*, dado que es lo que se desea en el comportamiento de los residuales y obtener la Gaussianidad del modelo. Además al calcular el parámetro  $\lambda$  que define la transformación de Box-Cox se aproxima a uno, entonces no es necesario aplicar la transformación a los datos. De manera análoga, este estudio se realizó para los demás subintervalo que poseen esta característica, es decir el cuarto, noveno, decimo, decimo primer y decimo segundo subintervalo los cuales no requieren de la transformación.

## RESIDUALES DE LOS MODELOS QUE REQUIEREN DE LA TRANSFORMACIÓN DE BOX-COX

Damos muestra en esta sección, de los residuales de los modelos que requieren la aplicación de la transformación de Box-Cox, esto es aquellos para los cuales la gráfica de cuantiles de los residuos no se aproxima a una línea recta sino más bien que posee comportamiento cóncavo o convexo. En este sentido, mostramos el primer subintervalo de la partición para hacer el estudio de la gráfica de cuantiles. Por un lado, mostramos los residuales originales y por el otro luego de la transformación de tal manera de ver los efectos que surte esta transformación a la hora de obtener la aproximación lineal buscada.

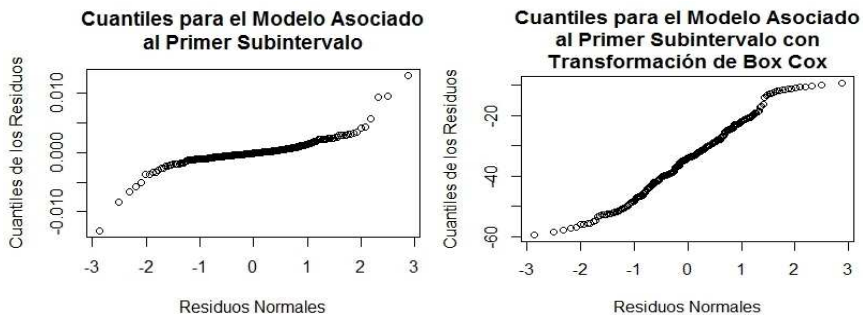


Figura 6: Gráficas de Cuantiles. Figura izquierda, el subintervalo original. Figura derecha, luego de la transformación Box-Cox.

En la Figura 6, se pueden notar las mejoras para el primer subin-

tervalo, observemos que en el gráfico de la izquierda el desarrollo de los cuantiles para el subintervalo original pasa de tener un comportamiento no lineal, a un comportamiento aproximadamente lineal, que es lo que se espera al hacer uso de la transformación Box-Cox. En este caso tenemos que el parámetro  $\lambda = -0,5520897$ , por lo que la aplicación de la transformación tiene sentido desde el punto de vista teórico.

La transformación Box-Cox sirve como herramienta para mejorar el comportamiento de la gráfica de cuantiles de los modelos en estudio, además si observamos y comparamos la evolución de los residuales normalizados, la ACF y los  $p$ -valores mediante el estadísticos de Ljung Box observamos que sobre todo los residuales normalizados muestran el aspecto y comportamiento aleatorio que se espera (Shumway y Stoffer, 2006; Cowpertwait y Metcalfe, 2008).

## PRONÓSTICOS DE LOS MODELOS ASOCIADOS A CADA SUBINTERVALO DE $[0,25; 1,45]$

Al elaborar pronósticos, el objetivo fundamental es predecir valores futuros de una serie de tiempo, basados en datos del presente. El objetivo del pronóstico en este trabajo es explorar que tan bien ajustan los modelos en estudio (Shumway y Stoffer, 2006). En este caso, lo que lo que se hizo fue predecir valores ya conocidos de la serie, es decir, puesto que conocemos 244 periodos de la tormenta y no podemos conocer más datos a futuro con precisión exacta, el procedimiento que empleamos consiste en trabajar con 234 periodos de los 244 y ajustar el modelo  $AR(1)$  asociado, para luego hacer el pronóstico de los 10 periodos que restan con una banda de confianza para la predicción hasta 95 %. Si los datos están dentro de esta banda de confianza, entonces esto es evidencia de un buen ajuste del modelo, además estos datos pronosticados deben aproximar a los valores originales que anteriormente fueron removidos.

En esta sección, se muestra la evidencia gráfica para algunos de los subintervalos en estudio. Para el caso de los subintervalos donde se usó la transformación de Box-Cox presentaremos ambos modelos para el primer subintervalo, el de los datos originales y al que se le aplicó la transformación de Box-Cox en conjunto con sus respecti-

vos pronósticos. El análisis para los subintervalos restantes que ya hemos mencionados se hace de la misma forma. Los modelos fueron obtenidos en base a la suma de los datos para el correspondiente intervalo con y sin la transformación de Box-Cox, según sea el caso y los residuos asociados al modelo  $AR(1)$ .

### Pronóstico para subintervalos con la transformación de Box-Cox

El pronóstico se muestra en la Figura 7, en donde al hacer la restricción de los 10 datos se tiene que el pronóstico realizado se encuentra dentro de la banda de confianza (líneas punteadas) hasta 95 %, se muestra el pronóstico anteriormente comentado, frente a los datos originales (líneas continuas), donde se evidencia una gran similitud entre el modelo encontrado y los datos originales, además los datos originales se hallan dentro de la banda de confianza considerada en este caso.

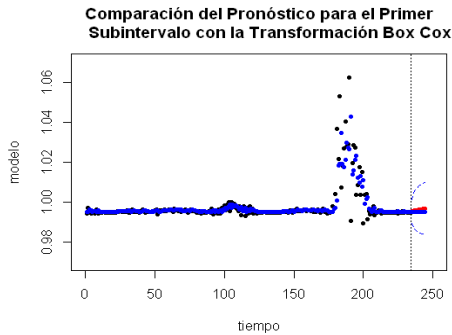


Figura 7: El subintervalo se muestra en líneas continuas, frente al modelo y su pronóstico para el modelo en puntos.

En general, para aquellos subintervalos donde se aplicó la transformación de Box-Cox se obtuvo que el pronóstico realizado para el respectivo modelo es bastante confiable en el sentido de que los 10 datos que hemos reproducido como se explicó anteriormente están dentro de la banda de confianza considerada casi en su totalidad, lo

que indica que estos pronósticos son capaces de *imitar* el comportamiento de los datos originales.

### Pronóstico para los Subintervalos que no Requieren Transformación Box-Cox

De la misma manera, que la sección anterior, procedemos a dar la prueba gráfica del pronóstico para los subintervalos que no poseen transformación de datos, en particular para el tercer subintervalo tenemos los siguientes pronósticos y además la comparación del modelo con los datos originales, de lo que se puede apreciar el buen ajuste del modelo con los datos de partida que hemos considerado. Además, como se observó en la Figura 5 los residuales para este modelo, al igual que para los demás subintervalos que ya se han mencionado que no requieren de la transformación, aproximan la línea recta que se requiere para obtener normalidad en los cuantiles. La Figura 8 a continuación muestra el pronóstico para el tercer subintervalo.

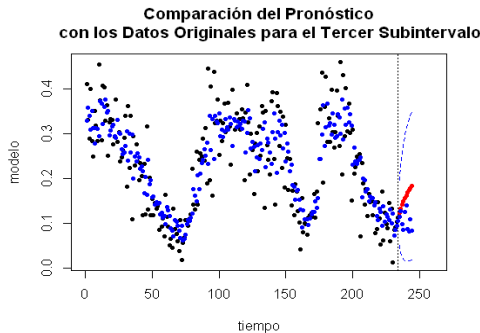


Figura 8: El subintervalo en línea continua, frente al modelo y su pronóstico en puntos.

## CONCLUSIONES

Con el fin de estudiar el espectro de los 244 periodos obtenidos en el Mar del Norte en 1999, se graficaron los espectros calculados a partir de los datos de la tormenta. Luego que se determinó el rango útil de frecuencias común a todos los espectros, se seleccionaron empíricamente valores a los extremos del rango, de modo que no se observaron características importantes fuera de este rango. Este forma el 97 % de la información total, y la energía total presente fuera del rango sólo representa el 3 %. Este rango útil, se dividió en subintervalos de forma tal que esta obedezca a capturar por cada subintervalo la mayor cantidad posible de cambios en la energía, esto con el objeto de explicar la evolución de la misma.

Después de este análisis, todos los subintervalos resultaron ser modelos autorregresivos, para los que se analizó el comportamiento de la ACF, PACF, los  $p$ -valores y el valor del AIC, más el estadístico de Ljung Box. Se determinó que cada subintervalo tiene asociado un modelo  $AR(1)$ . Para algunos de los subintervalos, fue necesario aplicar la transformación de Box-Cox, con el fin eliminar la no estacionaridad en la varianza y con esto lograr un mejor ajuste de los residuales asociado al modelo comparados con la distribución Normal. Para el otro grupo de subintervalos, se determinó mediante el calculo del parámetro  $\lambda$  de la Transformación Box-Cox era aproximadamente uno, por lo que no fue necesario aplicar dicha transformación, desde el punto de vista teórico.

Por último, se procedió a realizar el pronóstico para los modelos en cuestión con el fin de explorar el buen ajuste donde, el procedimiento usado fue truncar 10 datos de la serie de los 244 períodos y reproducir estos valores ya conocidos con el modelo, ya que hasta el momento no se tiene más información del fenómeno. Esta predicción se hizo con un 95 % de confianza, en algunos casos se obtuvo que el pronóstico fue capaz de imitar el comportamiento de los datos, este es el caso de aquellos períodos a los que se les aplicó la transformación de Box-Cox, mientras que, a los que no se les aplicó la transformación, a pesar de que el pronóstico queda dentro de la banda de confianza, no ajusta tan bien como en el caso anterior. Este análisis no es definitivo para explicar la tormenta, todavía falta construir modelos mas generales

que los ya analizados aquí para dar una mejor interpretación de la tormenta.

## Referencias

- Brockwell, P. J., y Davis, R. A. (2002). *Time introduction to time series and forecasting*. Springer.
- Cowpertwait, P. S., y Metcalfe, A. V. (2008). *Introductory time series analysis with r*. Springer.
- Lindgren, G. (2006). Lectures on stationary stochastic processes. a course for phd student in mathematical statistic and other fields. En *Lund university. centre for mathematical science. mathematical statistic*.
- Ochi, M. K. (2005). *Ocean waves. the stochastic approach*. Cambridge University.
- Ortega, J. (2001). Estudio de algunas propiedades del mar usando modelos aleatorios. *Boletín de la Asociación Matemática Venezolana*, VIII(2).
- Shumway, R. H., y Stoffer, D. S. (2006). *Time series analysis and its applications with r examples*. Springer.