



Universidad Central de Venezuela  
Facultad de Ciencias  
Escuela de Matemática

## **Análisis de máximos oleajes en datos del mar mediante el uso de series temporales**

Trabajo de Grado presentado ante la Ilustre Universidad Central de Venezuela  
por el Lic. Silfrido Gómez Polanco para optar al  
Titulo de Licenciado en Matemáticas Puras.

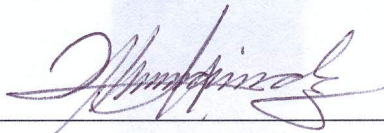
Caracas - Venezuela  
Mayo, 2014

Nosotros, los abajo firmantes, designados por la Universidad Central de Venezuela como integrantes del Jurado Examinador del Trabajo Especial de Grado titulado “Análisis de máximos oleajes en datos del mar mediante el uso de series temporales”, presentado por la **Lic. Silfrido Gómez Polanco**, titular de la Cédula de Identidad **17.589.849**, certificamos que este trabajo cumple con los requisitos exigidos por nuestra Magna Casa de Estudios para optar al título de **Licenciado en Matemática**.



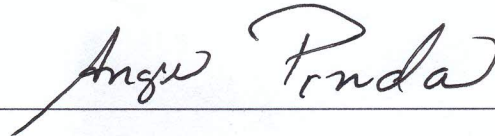
---

**Dra. Mairene Colina.**  
**Tutor**



---

**Dr. José B. Hernández**  
**Jurado**



---

**MSc. Angie Pineda.**  
**Jurado**

# Dedicatoria

A Dios todo poderoso, al universo, a mi madre (Ana Polanco), a mi sol (Wendy Gallardo) y a todas aquellas personas que siempre estuvieron allí para dar apoyo y una palmada en la espalda cuando más necesitaba ánimos.

# Agradecimientos

Primero que nada agradezco a Dios, a mis padres y a la energía universal por permitirme existir y vivir este momento. Agradezco a mi hermana Annkarys por ser tan comprensiva y paciente conmigo al igual que mis más sinceros agradecimientos a mi tutora Mairene Colina, pues élla siempre me brindo su apoyo y el impulso para culminar mi proyecto.

Gracias, Gracias e infinitas Gracias a todos los que me apoyaron.

# Contenido

<b>Contenido</b>	<b>iii</b>
<b>Introducción</b>	<b>1</b>
<b>1 Preliminares</b>	<b>4</b>
1.1 <b>Definiciones básicas</b> . . . . .	4
<b>2 Las olas y sus principios matemáticos</b>	<b>10</b>
2.1 <b>Características del Mar y las Olas</b> . . . . .	10
2.1.1 Anatomía de las olas. . . . .	10
2.1.2 Origen de las Olas . . . . .	12
2.1.3 Tipos de Olas . . . . .	13
2.1.4 El Mar en Tiempo y Espacio . . . . .	13
2.1.5 Altura Significativa de la Ola . . . . .	14
2.1.6 Propiedades del Mar . . . . .	15
2.2 <b>Modelos Estadísticos de las Olas</b> . . . . .	17
2.2.1 La Ola como proceso Gaussiano . . . . .	17
2.2.2 Modelos lineales de la Olas . . . . .	19
<b>3 Series de Tiempo</b>	<b>20</b>

---

<b>4</b>	<b>Análisis y Resultados</b>	<b>30</b>
4.1	Análisis de datos . . . . .	30
	<b>Bibliografía</b>	<b>48</b>

# Introducción

El hombre y el mar han mantenido una estrecha relación desde hace mucho tiempo, pues de manera indirecta, con el pasar de los tiempos y la evolución cognitiva del ser humano ha generado fuentes de trabajo, alimento, transporte y fuentes de recreación para gran parte de la población mundial.

En la últimas décadas el interés de la ciencia por el mar a mostrado un gran incremento en cuanto al estudio de sus características generales, comportamiento concerniente a las mareas y oleaje, severidades del mar, composición del mar, entre otros aspectos importantes que permiten a los investigadores tener un mayor acercamiento con dicho medio.

Ahora bien, si hacemos referencia al mar como objeto de estudio en cuanto al comportamiento de sus mareas y oleaje, es importante aclarar que la descripción de estos fenómenos históricamente no ha sido una labor sencilla para la ciencia específicamente para la matemática, pues dichos eventos estan conformados por una serie de variables que insiden directamente en su caracterización.

Un modelo matemático, aplicado para caracterizar este tipo de comportamientos está basado en la teoría de "Procesos Estocásticos", entendiendo como proceso estocástico

- Un conjunto de variables aleatorias  $X_t$  indexadas por un índice  $t$ , tal que  $t \in T, T \subseteq R$ .

Un ejemplo de este tipo de procesos son las llamadas Series Temporales las cuales son

---

una secuencia de datos, observaciones o valores, medidos en determinados momentos del tiempo, ordenados cronológicamente y, normalmente, espaciados entre sí de manera uniforme. Estas se usan para estudiar la relación causal entre diversas variables que cambian con el tiempo y se influyen entre sí. Desde el punto de vista probabilístico una serie temporal es una sucesión de variables aleatorias indexadas según un parámetro creciente con el tiempo.

Luego, una serie temporal empleada en datos recolectados del mar, muestra información importante sobre el comportamiento de su oleaje, corrientes, temperatura, entre otros; analizados en el tiempo, cabe destacar que dentro de los fines de este trabajo de investigación utilizaremos el oleaje y sus máximas alturas capturadas en distintos instantes del tiempo. Permitiendo esto último la predicción de la ola con mayor altura, es decir, los máximos oleajes.

Existen distintas maneras para la recolección de los datos que permiten esta clase de estudios, en este trabajo de investigación en particular se tomaron en cuenta datos recolectados de boyas localizadas en las costas del centro de California en una zona costera llamada: "Cabrillo Beach" y obtenidos mediante la extracción de la data en la página web The Coastal Data Information Program (CDIP). Es importante resaltar que se eligieron en esta localidad debido a la característica que presentan los datos, es decir, son datos sobre los cuales se puede aceptar el supuesto de gaussianidad en su distribución y que además es la hipótesis sobre la cual se han desarrollado las grandes teorías referentes al comportamiento del mar.

Para alcanzar el objetivo de este trabajo, el cual es realizar un análisis de máximas alturas en los datos del mar antes mencionados y determinar un modelo de series de tiempo para estas, es necesario tener dominio de los conceptos básicos de la teoría del modelaje del mar así como de la teoría de series temporales, es por ello que este trabajo está estructurado en

4 capítulos. El primer capítulo contiene definiciones previas que nos permiten tener claro las bases matemáticas, específicamente en el área de la estadística, en el segundo capítulo se exponen supuestos matemáticos bajo los cuales se describe el comportamiento del mar y sus alturas, en el tercer capítulo se dedicó a la explicación y definición de lo que es una serie de tiempo y los modelos aplicados durante el análisis de los datos y por último en el cuarto capítulo se mostraron los resultados obtenidos en el análisis mediante el uso de R-comander además de las gráficas obtenidas durante la investigación.

# Capítulo 1

## Preliminares

En este capítulo algunas definiciones básicas tales como: vector gaussiano, proceso gaussiano, función de distribución, variable aleatoria gaussiana, vector aleatorio entre otras serán presentadas de manera introductoria con la finalidad de facilitar la comprensión de este trabajo de investigación.

### 1.1 Definiciones básicas

A lo largo del trabajo consideraremos  $\Omega$  un espacio muestral y una  $\sigma$ -álgebra de subconjuntos de  $\Omega$  que denotaremos por  $\Lambda$ . Es habitual asignar a cada subconjunto  $A$  de  $\Lambda$  un número real  $P(A)$ , que llamaremos la probabilidad de  $A$ , de modo tal que se cumplan las siguientes condiciones:

- $P(A) \geq 0 \quad \forall A \subseteq \Lambda$ .
- $P(\Omega) = 1$  el evento cierto tiene probabilidad 1.

Si  $A_n \subseteq \Lambda$  para  $n = 1, 2, 3, \dots$  son disjuntos dos a dos, es decir, tales que  $A_i \cap A_j = \emptyset$  si  $i \neq j$ , entonces:

- $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ .

Una terna  $(\Omega, \Lambda, P)$  formada por un espacio muestral  $\Omega$ , una familia  $\Lambda$  de eventos y una probabilidad  $P$  se llama un *espacio de probabilidad*.

Ahora bien una vez conocido la definición de dicho espacio es muy importante recordar que una *variable aleatoria*  $X$  en un espacio de probabilidad es una variable de magnitud susceptible de tomar diversos valores con determinadas probabilidades, es decir  $X$  es una función real definida en el espacio muestral,  $\Omega$ , asociado a un experimento aleatorio.

También se hace pertinente resaltar los tipos de variables aleatorias con que se trabaja en probabilidad, entre las que se tienen:

- **Variable aleatoria discreta:** Es aquella cuyo recorrido es un conjunto discreto, es decir, sólo puede tomar un número finito numerable de valores.

*Ejemplo:* El número de componentes de una manada de lobos, puede ser 4, 5 ó 6 individuos, pero nunca 5,75 ó 5,87.

- **Variable aleatoria continua:** Es aquella cuyo recorrido es un conjunto no numerable, es decir, puede tomar un número infinito de valores en un intervalo dado.

*Ejemplo:* La densidad de una sustancia o altura de un árbol.

### Definición 1.1.1 (*función de distribución*)

Si  $X$  es una variable aleatoria, se define la función de distribución de  $X$  como la función  $F_X : \mathbb{R} \rightarrow [0, 1]$  tal que para cada  $y \in \mathbb{R}$ ,  $F_X(y) = P(\omega \in \Omega : X(\omega) \leq y) = P(X \leq y)$ .

nota: En algunos casos se utiliza la notación  $F$  en lugar de  $F_X$ .

Si  $X$  es una variable aleatoria discreta tal que  $X(\Omega) = (x_1, \dots, x_n, \dots)$ , entonces se define:

- $F_X(x) = \sum_{x_n \leq x} p_n$  donde  $p_n = P(X = x_n)$ .

Si  $X$  es continua:

- $F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt.$

Donde  $f$  es una función de densidad, es decir  $f(x) \geq 0 \forall x \in \mathbb{R}$  y  $\int f(x)dx = 1$ . También vale la pena recordar la definición de dos parámetros vinculados a las variables aleatorias que son de gran relevancia, los mismos son **La media o Esperanza de  $X$**  que está dada por:

$$\mu = E[X] = \int_{-\infty}^{\infty} x dP(x)$$

y **La varianza** que se expresa como:

$$\sigma^2 = var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 dP(x)$$

Luego dentro de los fines de esta investigación y como parte de su estructura lógica en el andamiaje teórico, es pertinente definir el significado de *variable aleatoria gaussiana* como parte previa a la definición de vector gaussiano.

**Definición 1.1.2** (*Variable aleatoria gaussiana*)

Sea  $X$  una variable aleatoria con función de densidad  $f$  igual a:

$$f(x) = e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\sigma}}$$

entonces  $X$  se denomina una *variable aleatoria gaussiana*, donde  $\mu, \sigma$  son la media y la varianza respectivamente.

**Nótese:** que cuando denotamos  $N(\mu, \sigma^2)$  significa que la variable aleatoria  $X$  es normal o gaussiana con promedio  $\mu$  y varianza  $\sigma^2$ .

**Definición 1.1.3** (*Vector aleatorio*)

Dado un espacio muestral  $\Omega$ , diremos que  $X = (X_1, X_2, \dots, X_k)$  es un vector aleatorio de dimensión  $K$  si cada una de sus componentes  $X_i : \Omega \rightarrow \mathbb{R}$ , para  $i = 1, \dots, k$  es una variable aleatoria. Notemos que el vector aleatorio  $X : \Omega \rightarrow \mathbb{R}^k$ .

En este punto se hace notable el nombrar dos definiciones importantes que forman parte de nuestro conjunto teórico como es el caso de la definición de la **esperanza** y la **matriz de covarianza de un vector**.

**Definición 1.1.4** (*Esperanza*)

La esperanza matemática de un vector  $X = (X_1, X_2, \dots, X_n)$  aleatorio se define como  $E(X) = (E[X_1], \dots, E[X_n]) = \mu_1, \dots, \mu_n$

**Definición 1.1.5** (*Matriz de Covarianza de un vector*)

Sea  $X = (X_1, X_2, \dots, X_n)$  un vector aleatorio donde cada  $X_i$  tiene varianza finita, entonces definimos la matriz de covarianza,  $\sum_{ij}$ , es la matriz cuyas entradas  $(i, j)$  es la covarianza

$$\sum_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$$

donde  $\mu_i = E(X_i)$ .

Por otro lado es importante destacar que los vectores aleatorios son del tipo discreto, continuo y mixto. Luego un **vector aleatorio** es **discreto o continuo** si y sólo si sus correspondientes componentes son todas variables aleatorias discretas o todas variables aleatorias continuas y será **mixto** si una o algunas de las componentes del vector son discretas y las otras continuas.

Un ejemplo de vector aleatorio continuo, y que nosotros usaremos con frecuencia, es un vector gaussiano cuya definición se muestra a continuación.

**Definición 1.1.6** (*Vector Gaussiano*)

Sea  $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$  un vector aleatorio. Se dice que  $X$  es un vector gaussiano si él tiene densidad normal multivariada esto es, si  $f(X) = \frac{1}{(2\pi|\sum_n|)^{\frac{n}{2}}} (\exp(-(x - \mu_n) \sum_n^{-1} (x - \mu_n)^t))$  donde  $\mu = (E[X_1], \dots, E[X_n])$  y  $\sum_k$  es la matriz de covarianza cuya entrada  $[(i, j)]$  está definida por  $\sum_k(i, j) = cov(X_i, X_j)$ .

Una manera de corroborar si un vector aleatorio  $X = (X_1, X_2, \dots, X_n)$  es un vector normal o gaussiano es verificando que toda combinación lineal de sus componentes es una variable aleatoria gaussiana.

Ahora, en este mismo orden de ideas se define como *proceso aleatorio* a toda experiencia que genera una secuencia de valores modelizables como variables aleatorias. Cada experiencia global tiene un posicionamiento, o sea un orden en la experiencia global. También se puede decir que una serie ordenada de experiencias individuales genera una experiencia global que se le denomina proceso aleatorio.

Seguidamente pasamos a definir proceso estocástico una vez revisadas las definiciones anteriores.

**Definición 1.1.7** (*Proceso Estocástico*) Un proceso estocástico  $X$  a tiempo continuo es un familia  $(X_t, t \in I) = (X_t(\omega), \omega \in \Omega, I \subseteq \mathbb{R})$ , de variables aleatorias definidas sobre un espacio  $\Omega$ , donde  $I$  es un intervalo, un conjunto finito o infinito numerable.

Un proceso estocástico es una función de dos variables:

- Para un instante de tiempo  $t$  fijo,  $X_t(\omega), \omega \in \Omega$ , es una variable aleatoria.
- Para un  $\omega$  fijo,  $X_t(\omega), t \in I$ , es una función del tiempo. Esta función es llamada trayectoria del proceso  $X$ .

**Definición 1.1.8** (*Proceso Gaussiano*)

Un proceso estocástico  $X_t, t \in I$  es llamado **gaussiano** si sus distribuciones finito-dimensionales son gaussianas, es decir, si para cada selección de  $0 \leq t_1 \leq t_2 \leq \dots \leq t_k$  el vector aleatorio  $Z = (X_{t_1}, \dots, X_{t_k}) \in \mathbb{R}^k$  tiene distribución normal multivariada.

Entonces, podemos definir un proceso  $X$  como un proceso estocástico que genera muestras en el tiempo, de manera tal, que para cada conjunto finito de índices  $t_1, \dots, t_k$  de un

conjunto  $\tau$ , tenemos que,  $\vec{X}_{t_1, \dots, t_k} = (X_{t_1}, \dots, X_{t_k})$  es un vector gaussiano, con un vector de medidas  $(\mu(t_1), \dots, \mu(t_k))$  y matriz de covarianza  $\Sigma = (E(X_{t_i} - \mu(t_i))(X_{t_j} - \mu(t_j)))$ .

Otro ejemplo de proceso gaussiano son las series de tiempo que definiremos con detalle en el capítulo 3.

Un tipo de proceso de interés son los procesos estacionarios.

**Definición 1.1.9** (*Proceso estacionario*)

El proceso  $(X_t)_{t \geq 0}$  es estacionario si para cualquier colección  $(t_i)_{i=1}^n$  de índices y cualquier  $h \in \mathbb{R}$  tal que  $t_i + h \geq 0$  se verifica que:

$$(X_{t_1}, \dots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_k+h})$$

.

# Capítulo 2

## Las olas y sus principios matemáticos

En este capítulo se exponen a continuación las definiciones matemáticas requeridas dentro de este proyecto de investigación que se encuentran asociadas al comportamiento del mar y las olas.

### 2.1 Características del Mar y las Olas

#### 2.1.1 Anatomía de las olas.

Las olas son ondas que se desplazan por la superficie de los mares y océanos, las cuales se ponen en movimiento mediante la acción del viento, éstas son el principal agente de modelado de las costas y están constituidas por :

- Línea de agua: es el nivel de la superficie del mar (si está en calma y si es perfectamente plano).
- Cresta: es el punto más alto de la ola, por encima de la línea de agua.
- Valle: es el punto más bajo de la ola, por debajo de la línea de agua.

- **Altura:** es la distancia vertical entre la cresta y el valle.
- **Longitud:** es la distancia horizontal entre las crestas sucesivas o depresiones.
- **Período:** es el tiempo que tarda una ola completa para pasar un punto en particular.
- **Frecuencia:** es el número de olas que pasan por un punto en un período de tiempo particular.
- **Amplitud:** es la mitad de la altura de las olas o la distancia desde la cresta hasta la línea de agua.
- **Profundidad:** es la distancia desde el fondo del océano hasta la línea de agua.
- **Dirección de propagación:** es la dirección en la que una ola viaja.

Esta descripción se puede apreciar gráficamente en la figura 2.1

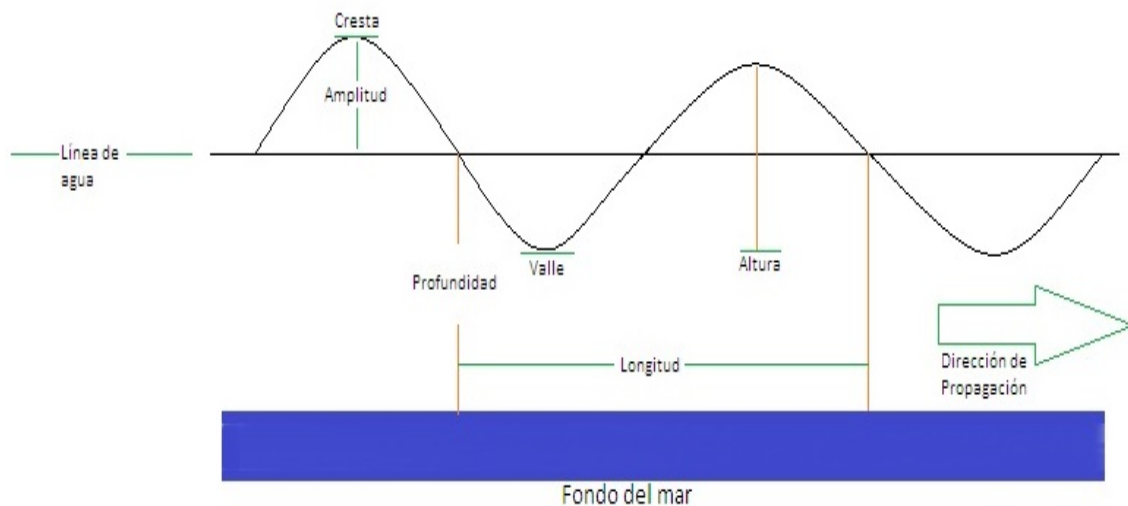


Figura 2.1: Anatomía de la Ola

### 2.1.2 Origen de las Olas

Las olas se originan por el rozamiento del viento sobre la superficie del mar. El viento ejerce un empuje sobre la superficie del mar que hace que se rice creando pequeñas ondas, que a su vez ofrecen más resistencia al viento lo que hace que se vayan transformando en ondulaciones cada vez mayores, es un proceso de "retro-alimentación", es decir cuanto mayor sea la onda mayor será el empuje que el viento podrá hacer sobre ella.

Las olas se originán por tanto dependiendo de la fuerza del viento, el tiempo durante el cual esté soplando y el área oceánica a la que afecte ese viento.

La gravedad terrestre también juega un papel ya que después de que el viento eleve la superficie del mar, la gravedad la vuelve a empujar hacia abajo, con lo que al subir y bajar se va acumulando una energía cinética asociada a la ola que forma.

Yendo más allá, nos podemos preguntar que origina el viento, pues bien, el viento se produce cuando existen variaciones de presión en la atmósfera, lo que se conoce como borrascas (área de bajas presiones) y anticiclones (área de altas presiones). Estos fenómenos son a su vez producidos por diferencias en la temperatura del aire, que a su vez depende del modo en que las radiaciones solares inciden sobre la atmósfera y los movimientos de rotación de la tierra.

Es importante resaltar que para los fines de esta investigación sólo se tomara en cuenta el origen de las olas mediante la insidencia del viento y **los maximos** (definición que más adelante será expuesta) alcanzados por dicho fenómeno dentro del registro de olas con que se trabaja.

### 2.1.3 Tipos de Olas

Las olas, cuando llegan a tierra firme pueden ser de tres tipos, **Mar de fondo** o **swell**, **olas de viento** y **Mixtas**.

El **Mar de fondo** o **swell** son olas que originariamente se crearon por incidencia del viento en alguna área del océano, pero que a medida que viajan por el mar dejan de estar afectadas por el viento que las originó, se alisan y se ordenan, pero mantienen su tamaño.

Las **olas de viento** son las que llegan a la costa sin tiempo para ordenarse. Esto es así porque se originan cerca de la costa por vientos que la afectan desde que se crean mar adentro hasta que rompen en la playa. Son por lo general olas de tamaño pequeño.

Las olas **Mixtas** son olas que llegan a la costa y resultan de ambos tipos de olas mencionados en la parte de arriba, es decir que se asocia un mar de fondo con un fuerte viento en la costa dando lugar a un tipo de ola **Mixta**.

### 2.1.4 El Mar en Tiempo y Espacio

Supondremos que la superficie del Mar puede ser modelada como una "superficie aleatoria" que evoluciona en el tiempo, es decir, como un proceso aleatorio que depende de la posición en el espacio  $x$ , del tiempo  $t$ , y por supuesto, de un parámetro aleatorio  $\omega$  que pertenece a un espacio de probabilidad  $(\Omega, \Lambda, P)$  :  $M(t, \mathbf{x}, \omega)$ . Por lo general se omite el parámetro  $\omega$ .

Los datos estadísticos del oleaje que usaremos han sido obtenidos de boyas flotantes (ubicadas en costas de fondos oceánico). Esto significa que describen el oleaje como un proceso en el tiempo en un punto fijo. Usamos la notación  $M(t)$  para denotar la altura



Figura 2.2: Anatomía de la Ola

del nivel del mar en un punto fijo como función del tiempo  $t$ . En este caso decimos que depende del tiempo y lo que obtenemos es una serie temporal. Dicha serie puede denotarse de la siguiente forma:

$$M(t) = M(t, x) \text{ con } x \text{ fijo.}$$

### 2.1.5 Altura Significativa de la Ola

La manera más simple para describir la severidad del estado del mar es por alguna medida de variabilidad de la elevación de la ola  $M$  por arriba y debajo del nivel medio del mar. La medida más sencilla es la variación estándar de la elevación  $M$ , calculada como la desviación estándar observada en una larga secuencia de valores de las muestras  $z_1, \dots, z_N$ , donde  $N$  es el número de registros.

Para series de datos suficientemente largos, se trata de una estimación precisa de la ver-

dadera desviación estándar de la elevación del mar, es decir,  $\sqrt{V(M(t))}$ . Es importante señalar que ésta medida de la severidad del mar no depende ni del tiempo ni de la ubicación en el espacio, es decir,

$$V(M(t)) = V(M(x)).$$

La medida más importante de la severidad del mar es la altura significativa de la ola. Esta medida indica la altura de las olas más severas que puede encontrarse durante un plazo razonable de tiempo.

### Definición de Altura Significativa de la Ola

La altura Significativa de una Ola del estado del modelo del mar es definido como:

$$H_\sigma = 4\sigma,$$

donde

$$\sigma = \sqrt{\sigma^2} = \sqrt{V(M(t))}$$

es la desviación estándar del proceso del nivel del mar.

La altura Significativa de una Ola de un estado del mar observado es

$$S = \sqrt{\frac{1}{N} \sum_{k=1}^N (z_k - \bar{z})^2}$$

donde  $z_1, \dots, z_N$  es la secuencia de las mediciones independientes del nivel del mar y  $\bar{z}$  es el nivel medio del mar y  $N$  es el número de registros.

## 2.1.6 Propiedades del Mar

Cuando se quiere obtener un modelo manejable que nos permita hacer un estudio de las propiedades del mar es necesario tomar en cuenta las siguientes hipótesis:

1. Permanencia del tiempo de las condiciones del mar: Supondremos que el proceso tomado como modelo es de tipo estacionario. Es decir, la distribución de  $M(t + h)$  es la misma para cualquier valor de  $h$ , particularmente siempre es igual para  $M(0)$ . Más aún, para cualquier valor  $n$  y cualesquiera instantes de tiempo  $t_1, \dots, t_n$  la distribución del vector  $M(t_1 + h), \dots, M(t_n + h)$  es independiente del valor de  $h$ .
2. Proceso centrado: Supondremos que el nivel medio del mar es 0 y se medirán las variaciones con respecto al mismo. Es decir, se considera el proceso como  $E(M(t)) = 0$ .
3. Las trayectorias del Proceso  $M$  sean continuas:  $M$  es una función sobre el espacio producto  $[0, \infty) \times \Omega$  con la propiedad que para cada  $t \in [0, \infty)$  fijo  $M(t, \diamond)$  *medible*. Si fijamos  $\omega \in \Omega$  obtenemos una función.

$$M(\diamond, \omega) : [0, \infty) \rightarrow \mathbb{R},$$

que se conoce como *trayectoria del proceso*. Pedimos que para casi todo  $\omega \in \Omega$  (que tenga probabilidad 1) esta función sea continua.

4. El proceso sea ergódico: Para no entrar en tecnicismos de esta hipótesis, es la que permite sustituir los valores esperados (teóricos) por promedios temporales (empíricos).

$$E(M(t)) = \int_{\Omega} M(t, \omega) dP(\omega) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t M(u) du.$$

$$Cov(M(t), M(t+h)) = \int_{\Omega} M(t, \omega) M(t+h, \omega) dP(\omega) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t M(u) M(u+h) du.$$

5. Gaussianidad: Esto indica que la distribución de la altura de la ola en un punto fijado y en un instante de tiempo  $t$  tiene distribución gaussiana.

La hipótesis de Gaussianidad nos dice que para cualquier valor de  $n$  y cualquiera que sea el instante de tiempo  $t_1, \dots, t_n$  la distribución del vector  $M(t_1 + h), \dots, M(t_n + h)$

tiene *densidad gaussiana*

$$f_{t_1, \dots, t_n}(u_1, \dots, u_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} u' \Sigma^{-1} u\right)$$

donde  $u = (u_1, \dots, u_n)$

6. La función de Covarianza debe satisfacer lo siguiente:

$$r(s, t) = \text{Cov}(M(s), M(t)) = E(M(s)M(t)) = r(|s - t|).$$

En particular si  $s = t$

$$r(s, s) = \text{Var}(M(s)) = E(M(s)^2) = r(0).$$

La función de covarianza  $r$  es par y por tanto, si es diferenciable en 0, la derivada debe ser nula. Más aún, si  $r$  tiene dos derivadas en el origen la segunda derivada debe ser negativa :  $r''(0) < 0$ .

## 2.2 Modelos Estadísticos de las Olas

### 2.2.1 La Ola como proceso Gaussiano

Sea  $\eta$  un perfil de olas de un tiempo fijo  $t$ . Aquí  $\eta$  es una variable aleatoria en el espacio  $(-\infty, \infty)$ .

Suponemos que  $\eta$  es la suma de un número grande de componentes de  $X$ , esto es

$$\eta = x_1 + \dots + x_n,$$

donde los  $x_i$  son variables aleatorias estadísticamente independientes, con la misma distribución de probabilidad.

Sea el valor medio de  $x_i$  cero ( $E[x_i] = 0$ ) y su varianza  $Var[x_i] = \sigma^2$ .

Dado que las  $x_i$  son variables aleatorias y por el Teorema Central de Límite la distribución  $\eta$  tiene media 0 y varianza  $\sigma^2$ .

Podemos por lo tanto estandarizar la variable  $\eta$  y escribimos la nueva variable estandarizada  $Z$ , de modo que

$$Z = \frac{\eta}{\sqrt{n}\sigma} = \sum_{j=1}^n \frac{X_j}{\sqrt{n}\sigma}.$$

Ahora bien  $\phi_x(t)$  la función característica de  $X$ . Dado que las  $X_j$ ,  $j = 1, \dots, n$  son variables aleatorias independientes se sigue que la función característica de  $Z$  viene dada por

$$\phi_z(t) = [\phi_x(\frac{t}{\sqrt{n}\sigma})]^n.$$

Por otra parte, la función característica en general se puede expandir como

$$\phi_x = 1 + itE[x] - \frac{t^2}{2}E[X^2] + \dots$$

Dado que para una variable estandarizada,  $E[x] = 0$  y  $E[X^2] = 1$  entonces podemos escribir  $\phi_x(\frac{t}{\sqrt{n}\sigma})$  como sigue

$$\phi_x(\frac{t}{\sqrt{n}\sigma}) = 1 - \frac{t^2}{2n} + O(\frac{t^2}{n})$$

donde  $O(\frac{t^2}{n})$  indica que el resto del desarrollo es una función que tiende a cero de la misma manera que  $\frac{t^2}{n}$  tiende a cero. Así, tenemos

$$\phi_z(t) = [1 - \frac{t^2}{2n} + O(\frac{t^2}{n})]^n$$

haciendo  $n \rightarrow \infty$ , obtenemos

$$\lim_{n \rightarrow \infty} \phi_z(t) = \lim_{n \rightarrow \infty} [1 - \frac{t^2}{2n} + O(\frac{t^2}{n})]^n,$$

de donde

$$\phi_z(t) = \exp\left(-\frac{t^2}{2}\right)$$

la cual es la función característica de una distribución  $N(0,1)$ . Dado que las funciones características son únicas esto implica que la variable aleatoria  $Z$  tiene distribución  $N(0,1)$ .

Así el proceso  $\eta$  tiene distribución normal de media 0 y varianza  $n\sigma^2$ , por consiguiente podemos decir que el proceso que describe las olas es un proceso *Gaussiano*.

### 2.2.2 Modelos lineales de la Olas

Usaremos  $M(t)$  como la descripción de una ola aleatoria en función del tiempo. La manera matemática más sencilla de sintetizar una ola aleatoria es como la suma de un número finito de elementos armónicos de olas con diferentes frecuencias angulares.

$$M(t) = \sum_{k=1}^{\infty} A_k \cos(W_k t + \varepsilon_k), \quad (2.1)$$

donde  $A_k$ ,  $k = 1, 2, \dots$  son amplitudes elegidas aleatoriamente con algunas distribuciones estadísticas, y  $\varepsilon_k$  son fases aleatorias. La frecuencia  $W_k$  son en esta representación frecuencias constantes, distribuidas sobre el rango entero de frecuencia de las olas. Los términos son asumidos por separados en  $M(t)$  para ser estadísticamente independientes de cada interacción.

Una forma alternativa para representar la expresión (2.1) mencionada anteriormente es la siguiente:

$$M(t) = \sum_{k=1}^{\infty} X_k \cos(W_k t) + Y_k \sin(W_k t),$$

donde  $X_k$  y  $Y_k$  son cantidades aleatorias, tal que  $A_k = \sqrt{X_k^2 + Y_k^2}$  y  $\cos(\varepsilon_k) = \frac{X_k}{A_k}$ ,  $\sin(\varepsilon_k) = -\frac{Y_k}{A_k}$ .

# Capítulo 3

## Series de Tiempo

De capítulos anteriores conocemos la definición de un proceso gaussiano y también sabemos que un ejemplo particular de esta clase de proceso es una serie temporal. En este capítulo se presentan definiciones matemáticas importantes que contribuyen a la comprensión de herramientas implementadas durante el análisis de los datos, y que además forman parte del análisis de una serie temporal que en nuestro caso se muestra más adelante representando las alturas máximas de olas en cierto período de tiempo .

A continuación definiremos lo que es una serie temporal como punto principal en la comprensión de dicho apartado.

### **Definición 3.0.1** (*serie de tiempo*)

Una *serie de tiempo* es un conjunto de observaciones  $x_t$ , cada una registrada a un tiempo específico  $t$ .

Observando la serie temporal de los datos del oleaje del mar determinamos patrones periódicos que nos ayudaron en el ajuste del modelo a implementar durante el análisis.

En este trabajo de investigación particularmente analizaremos la serie temporal de datos provenientes del oleaje del mar de las costas californianas.

A continuación presentaremos la definición de modelo de serie temporal, pues mantiene una hilaridad en los contenidos procedimentales desde el punto de vista cognitivo en cuanto a su comprensión como proceso matemático.

**Definición 3.0.2** (*modelo de series de tiempo*)

Un *modelo de series de tiempo* para los datos observados  $x_t$  es una especificación de la distribución conjunta (o posiblemente sólo de medias y covarianzas) de una sucesión de variables aleatorias  $X_t$  de las cuales  $x_t$  es una realización.

Una vez mostradas estas dos definiciones se procede con la conceptualización de una herramienta importante en el proceso del análisis y resultados, como lo es la función de auto correlación, pues ella mide la relación estadística entre los datos observados.

Generalmente esta función es aplicada sobre los datos residuales de los distintos modelos implementados sobre los datos como parte de la corroboración en el proceso de ajuste.

**Definición 3.0.3** (*Función de Autocorrelación*)

Dado un proceso  $X_t$ ,  $t \in T$  definimos la *función de autocorrelación (ACF)* como

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}.$$

Donde  $\gamma(s, t) = Cov(X_s, X_t)$ .

La (ACF) mide la predictibilidad lineal de una serie de tiempo en un tiempo  $t$ , digamos  $x_t$  usando sólo el valor  $x_s$ . Es fácil de demostrar que  $-1 \leq \rho(s, t) \leq 1$  usando la desigualdad de Cauchy-Schwarz. Si podemos predecir  $x_t$  exactamente de  $x_s$  a través de la relación lineal  $x_t = \beta_0 + \beta_1 x_s$  entonces la correlación será 1 cuando  $\beta_1 \geq 0$  y  $-1$  cuando  $\beta_1 \leq 0$ .

*Nótese* que la desigualdad de Cauchy-Schwarz implica

$$|\gamma(s, t)|^2 \leq \gamma(s, s)\gamma(t, t).$$

Sabemos ya de las dos definiciones anteriores que un modelo de series temporales nos permite asociar los datos con los que se trabajan a una posible distribución conjunta; además mediante la función de autocorrelación se obtiene información sobre que tan predecible es la relación lineal de los datos y así de esta manera se continúa buscando patrones de comportamientos similares entre los mismos. A continuación se presenta la definición de modelo autoregresivo la cual pretende orientar al lector en la comprensión estructurada y como primer modelo de prueba en el análisis de los datos.

**Definición 3.0.4** (*Ruido Blanco*)

Una manera sencilla de generar series de tiempo puede ser considerando una sucesión de variables aleatorias no-correlacionadas,  $\omega_t$  con media 0 y varianza  $\sigma_\omega^2$ . Las series de tiempo generadas de esta manera son usadas como modelos para ruido en aplicaciones de ingeniería, donde ellas son llamadas *ruido blanco*, denotaremos este proceso como  $\omega_t \sim \omega n(0, \sigma_\omega^2)$ . La designación blanco se origina de la analogía con luz blanca e indica que todos los posibles periodos de oscilación están presente con igual intensidad.

**Definición 3.0.5** (*modelo autoregresivo de orden p*)

Un *modelo autoregresivo de orden p*, abreviado **AR(p)** es de la forma

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_{t-p} x_p + \omega_t, \quad (3.1)$$

donde  $x_t$  es estacionario,  $\phi_1, \phi_2, \dots, \phi_p$  son constantes ( $\phi_p \neq 0$ ). A menos que se declare lo contrario, se asume que  $\omega_t$  es un ruido blanco gaussiano de media 0 y varianza  $\sigma_w^2$ . La media de  $x_t$  en (1) es cero. Si la media de  $x_t$  no es cero, reemplazamos  $x_t - \mu$  en (1), es decir

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \dots + \phi_{t-p}(x_p - \mu) + \omega_t$$

o escribimos

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_{t-p} x_p + \omega_t, \quad (3.2)$$

donde  $\alpha = \mu(1 - \phi_1 + \dots + \phi_p)$ .

Dicho modelo es quien nos permite corroborar, como primera opción, la independencia o no de los datos de salida en relación a sus valores anteriores, es decir si se aplica un autoregresivo de orden cero (0) y este es el indicado en el proceso entonces sólo el ruido blanco o termino de error es quien contribuye en dicho proceso.

**Definición 3.0.6** (*modelo de promedio móvil de orden q*)

Un *modelo de promedio móvil de orden q* o modelo **MA(q)** se define como

$$x_t = \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \theta_p \omega_{t-p},$$

donde  $\theta_1, \theta_2, \dots, \theta_q, \in \mathbb{R}$  con ( $\theta_q \neq 0$ ) son parámetros. El ruido  $\omega_t$  se asume como ruido blanco gaussiano.

Este modelo nos permite suavizar la forma de la serie temporal, es decir, promedia los datos de manera que cada  $q$  elementos son un dato del número total y disminuye la cantidad entre los que se divide la suma de todos. Esto permite que la longitud de arco de la curva que modelan los datos en la serie temporal sea más amplia.

La combinación de ambos modelos descritos anteriormente define el modelo ARMA, el cual se menciona como continuación de este apartado.

Una definición de utilidad en lo que sigue es la de *proceso estacionario*

**Definición 3.0.7** (*Proceso estacionario*)

El proceso  $(X_t)_{t \geq 0}$  es estacionario si para cualquier colección  $(t_i)_{i=1}^n$  de índices y cualquier  $h \in \mathbb{R}$  tal que  $t_i + h \geq 0$  se verifica que:

$$(X_{t_1}, \dots, X_{t_k}) =^d (X_{t_1+h}, \dots, X_{t_k+h})$$

**Definición 3.0.8** (*ARMA(p,q)*)

Una serie de tiempo  $x_t, t = 0, \pm 1, \pm 2, \dots$  es **ARMA(p,q)** si es estacionario y

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_{t-p} x_p + \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \omega_{t-p} \theta_p,$$

con constantes  $(\phi_k \neq 0, k = 1, \dots, p)$  y  $(\theta_j \neq 0, j = 1, \dots, q)$  y  $\sigma_w^2 > 0$ . Los parámetros  $p$  y  $q$  son llamados ordenes autoregresivos y de promedio móvil respectivamente. Si  $x_t$  tiene media  $\mu$  distinta de cero, hacemos  $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$  y escribimos el modelo como

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_{t-p} x_p + \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \omega_{t-p} \theta_p.$$

A menos que se declare lo contrario  $(\omega_t, t = 0, \pm 1, \pm 2, \dots)$  es una sucesión de ruido blanco gaussiano.

Se debe destacar que una herramienta que nos permite chequear si el modelo en cuestión funciona para nuestro ajuste es la función de autocorrelación parcial y la función de autocorrelación.

**Definición 3.0.9** (*función de autocorrelación parcial*)

La *función de autocorrelación parcial* (PACF) de un proceso estacionario  $x_t$  denotada  $\phi_{hh}$  para  $(h = 1, 2, \dots)$  es

$$\phi_{11} = \text{corr}(x_1, x_0) = \rho(1)$$

y

$$\phi_{hh} = \text{corr}(x_h - x_h^{h-1}, x_0 - x_0^{h-1}) \text{ para } h \geq 2.$$

Tanto  $(x_h - x_h^{h-1})$  como  $(x_0 - x_0^{h-1})$  son no-correlacionados con  $(x_1, x_2, \dots, x_{h-1})$ . Por estacionariedad, la PACF  $\phi_{hh}$  es la correlación entre  $x_t$  y  $x_{t-h}$  con la dependencia lineal  $(x_{t-1}, \dots, x_{t-(h-1)})$  removida en cada uno.

Si el proceso  $x_t$  es gaussiano, entonces  $\phi_{hh} = \text{corr}(x_t, x_{t-h} \mid x_{t-1}, \dots, x_{t-(h-1)})$ . Esto es  $\phi_{hh}$  es el coeficiente de correlación entre  $x_t$  y  $x_{t-h}$  en la distribución binaria  $(x_t, x_{t-h})$  condicionada por  $(x_{t-1}, \dots, x_{t-(h-1)})$ .

Siguiendo este mismo orden se hace de gran importancia aclarar dos definiciones previas al concepto de (ARIMA), como lo son: operador de cambio y diferencia de orden  $d$ , para que de esta manera el lector tenga una idea más íntegra de dicho modelo al momento de su implementación como herramienta en el análisis de series temporales.

Ambas definiciones se presentaran consecutivamente, pues ellas forman parte en el desarrollo matemático del modelo (ARIMA)

**Definición 3.0.10** (*operador de cambio*)

Definimos el *operador de cambio* por

$$Bx_t = x_{t-1},$$

y extendemos a la potencia  $B^2x_t = B(Bx_t) = B_{t-1} = x_{t-2}$  y así sucesivamente. Entonces

$$B^k x_t = x_{t-k}.$$

Es claro entonces que podemos reescribir  $\nabla x_t = x_t - x_{t-1}$  como

$$\nabla x_t = (1 - B)x_t.$$

**Definición 3.0.11** (*diferencias de orden  $d$* )

Las *diferencias de orden  $d$*  son definidas como

$$\nabla^d = (1 - B)^d$$

donde podemos desarrollar el operador  $(1 - B)^d$  algebraicamente para evaluar valores superiores enteros de  $d$ .

Ahora se puede definir más explícitamente el modelo siguiente ya que parte de sus componentes en la construcción del mismo se aclaran en estas líneas anteriores.

**Definición 3.0.12** (*ARIMA( $p, d, q$ )*)

Un proceso  $x_t$  se dice que es *ARIMA( $p, d, q$ )* si

$$\nabla^d x_t = (1 - B)^d x_t$$

es *ARMA( $p, q$ )*. En general, escribiremos el modelo como

$$\phi(B)(1 - B)^d x_t = \theta(B)\omega_t,$$

donde

$$\phi(B) = 1 + \phi_1(B) + \phi_2(B^2) + \dots + \phi_p(B^p),$$

$$\phi_i(B^k)(x_t) = \phi_i x_{t-k}$$

y  $\phi_i \in \mathbb{R}$  además

$$\theta_B = 1 + \theta_1(B) + \theta_2(B^2) + \dots + \theta_p(B^p),$$

$$\theta_i(B^k)(x_t) = \theta_i x_{t-k}$$

y  $\theta_i \in \mathbb{R}$ .

Si  $E(\nabla^d x_t) = \mu$  entonces escribiremos el modelo como

$$\phi(B)(1 - B)^d x_t = \alpha + \theta(B)\omega_t$$

, donde  $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$ .

Se pueden realizar modificaciones a los modelos ARIMA para explicar el comportamiento estacional y no estacionario. A menudo, la dependencia en el pasado tiende a ocurrir mas fuertemente en múltiplos de algún paso estacional subyacente  $s$ . Por ejemplo, con datos económicos mensuales, existe un fuerte componente anual ocurrida en pasos que son múltiplos de  $s = 12$ , debido a la fuerte conexión de todas las actividades en el año calendario. Los datos tomados trimestralmente exhiben un período repetitivo anual con  $s = 4$  trimestres. Fenómenos naturales tal como la temperatura, oleaje, etc., también tienen una componente fuerte correspondiente a la estación. Debido a esto, es apropiado introducir polinomios autoregresivos y de promedio móvil que se identifiquen con las fases estacionales. El modelo autoregresivo de promedio móvil puramente estacional ARMA(P,Q) $_s$  tiene entonces la forma:

$$\Phi_P(B^s)x_t = \Theta_Q(B^s)\omega_t$$

donde los operadores  $\Phi_P(B^s)$  y  $\Theta_Q(B^s)$  satisfacen la siguiente igualdad:

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}.$$

y

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs}.$$

son el operador autoregresivo estacional y el operador de promedio móvil estacional de ordenes  $P$  y  $Q$  respectivamente con periodo estacional  $s$ . En general, podemos combinar los operadores estacional y no estacional en un modelo autoregresivo de promedio móvil estacional multiplicativo, denotado por  $ARMA(p, q) \times (P, Q)_s$  y escribir

$$\Phi_P(B^s)\phi(B)x_t = \Theta_Q(B^s)\theta(b)\omega_t$$

como el modelo general. En la adaptación de estos modelos, centrándonos en los componentes autoregresivos de promedio móvil estacional primero generalmente conduce a resultados más satisfactorios.

La estacionalidad no estacionaria puede ocurrir, por ejemplo, cuando el proceso es aproximadamente periódico en la estación. Por ejemplo, con promedios de temperatura mensual a lo largo de un año, cada enero será aproximadamente igual, cada febrero será aproximadamente igual, y así sucesivamente. En este caso, podemos pensar en la temperatura promedio mensual  $x_t$  como un modelo

$$x_t = S_t + \omega_t$$

donde  $S_t$  es una componente estacional que varía lentamente de un año a otro, de acuerdo a un camino aleatorio

$$S_t = S_{t-12} + v_t.$$

En este modelo,  $\omega_t$  y  $v_t$  son ruidos blancos no correlacionados. Si restamos el efecto de los años sucesivos el uno del otro, encontramos, que

$$(1 - B^{12})x_t = x_t - x_{t-12} = v_t + \omega_t - \omega_{t-12}$$

Este modelo es un  $MA(1)_{12}$  estacionario. En general, la diferenciación estacional puede indicar cuando la ACF decae lentamente en múltiplos de alguna estación  $s$ , pero que es insignificante entre estos periodos. Entonces, una diferencia estacional de orden  $D$  se define como

$$\nabla_s^D x_t = (1 - B^s)^D x_t$$

donde  $D = 1, 2, \dots$  toma valores enteros. Normalmente,  $D = 1$  es suficiente para obtener estacionaridad estacional. Incorporando estas ideas al modelo general nos da la siguiente definición:

**Definición 3.0.13** (*modelo autoregresivo de promedio móvil estacional*)

Un *modelo autoregresivo de promedio móvil estacional* multiplicativo o modelo **SARIMA**, de Box y Jenkins (1970) está dado por

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d x_t = \alpha + \Theta_Q(B^s)\theta(b)\omega_t$$

donde  $\omega_t$  es el proceso de ruido blanco gaussiano usual. El modelo general se denota como **ARIMA(p,d,q) × (P,D,Q)<sub>s</sub>**. Las componentes autoregresivas y de promedio móvil ordinarias están representadas por los polinomios  $\phi_p$  y  $\theta_q$  de ordenes p y q respectivamente, y las componentes autoregresivas y de promedio móvil estacional por  $\Phi_P(B^s)$  y  $\Theta_Q(B^s)$  de ordenes P y Q y las componentes de diferencias ordinaria y estacional por  $\nabla^d = (1 - B)^d$  y  $\nabla_s^D = (1 - B^s)^D$ .

Vale la pena destacar que el último tipo de modelo de series temporales fue el que mejor se ajustó a los datos estudiados en este trabajo de investigación.

# Capítulo 4

## Análisis y Resultados

### 4.1 Análisis de datos

Para el cálculo de máximos oleajes como se mencionó en el capítulo 2, los datos deben ser gaussianos. La página **CDIP** contiene diferentes registros de oleajes de las costas norteamericanas. Nosotros en este trabajo de investigación consideramos datos de las costas californianas específicamente hacia el centro en una locación llamada Cabrillo Beach y bajo el supuesto de gaussianidad de dichos datos se estimó el máximo oleaje usando la teoría expuesta en las preliminares y en el capítulo 3.

La figura 4.1 nos muestra la serie temporal de los datos (mar), en la cual para el eje de las abscisas tenemos el tiempo marcado desde el primer dato hasta el último y en el eje de la ordenada las máximas alturas alcanzadas para cada instante determinado. La serie inicia específicamente el 03/02/2010 (3 de Febrero del 2010) alcanzando una altura máxima de 2.49 metros y finaliza 08/07/2012 (8 de Julio del 2012) con una altura máxima de 2.34 metros.

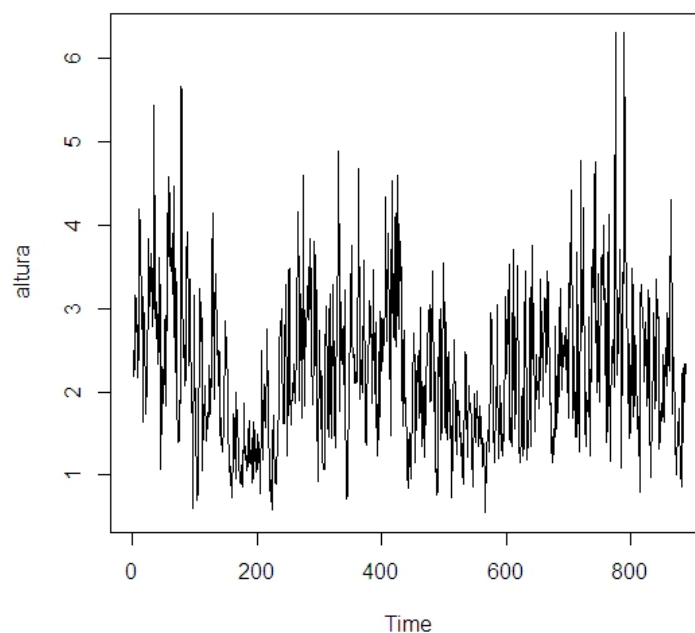


Figura 4.1: Serie Temporal

Además podemos apreciar un patrón de periodicidad así como oleajes que alcanzan alturas superiores a los 5 metros y oleajes con alturas inferiores a 1 metro.

Para comprender más en detalle las características de los datos, hicimos primero un gráfico de caja de los datos originales.

Como se puede apreciar en la figura 4.2, la mediana alcanza un valor de 2.21 metros, las alturas se encuentran concentradas entre el percentil 25 con un valor de 1,60 metros y el percentil 75 con un valor de 2,84 metros. Notamos alturas mínimas de 0,55 metros y máximas de 5 metros con algunas posibles alturas atípicas superiores a los 5 metros, es importante resaltar que mediante este gráfico no se visualiza fácilmente la media, y por esta razón se

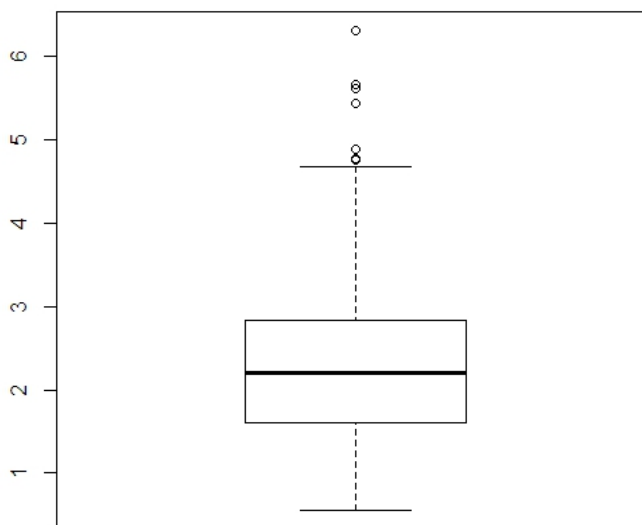


Figura 4.2: gráfico de caixa

calculó numericamente mostrando un valor de 2.27 metros de altura. Seguidamente presentamos un histograma de frecuencia de dichos datos donde podremos notar la distribución los mismos y el dato más repetido, es decir, la moda. Nótese que en el eje horizontal la escala correspondiente es con respecto a las alturas maximas alcanzadas y el eje vertical se asocia directamente con la frecuencia.

Para continuar con la descripción de la distribución de los datos originales, seguidamente realizamos un histograma de frecuencia figura 4.3 e indicamos los valores de la media y la varianza:

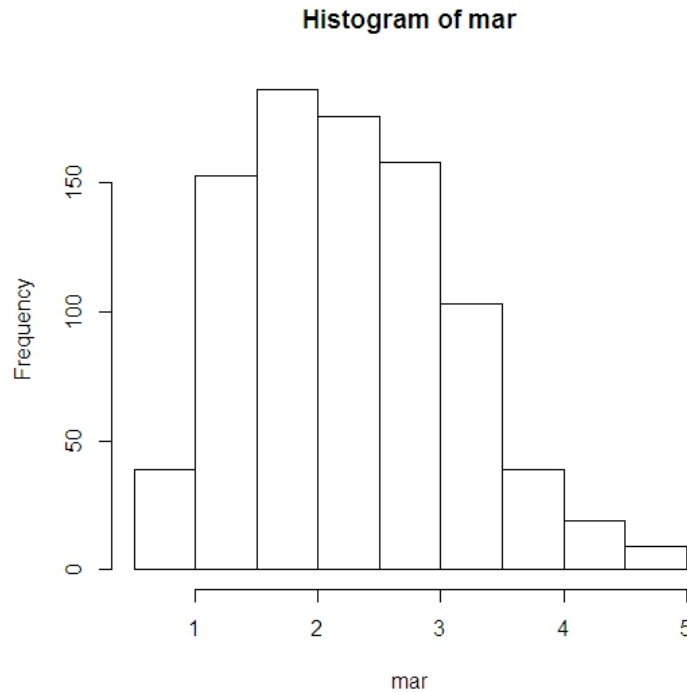


Figura 4.3: Histograma

Como se puede apreciar en el gráfico de barra del histograma de frecuencia figura 4.3, se verifica la forma semi acampanada que nos podría hacer suponer que los datos tiene distribución gaussiana, el valor de la media de dichos datos es  $\mu = 2.27$  y su varianza  $\sigma^2 = 0.78$ .

Nótese que el principio de gaussianidad supuesto sobre los datos originales se mantiene sobre los maximos oleajes y además al extraer los datos atípicos obtendríamos una campana más centrada.

Siguiendo este mismo orden de ideas, en lo que respecta al tratamiento de la serie temporal vinculada a esta investigación y como producto de la volatilidad de los datos, es decir, de pasar de tener alturas pequeñas entre 1 y 2 metros, a alturas de más de 6 metros es difícil

apreciar y determinar el patrón de comportamiento de estos. Por ello se procede a filtrar los mismos mediante la función logaritmo. A continuación se presentará un gráfico de caja (boxplot) de los datos una vez aplicada la función logaritmo:

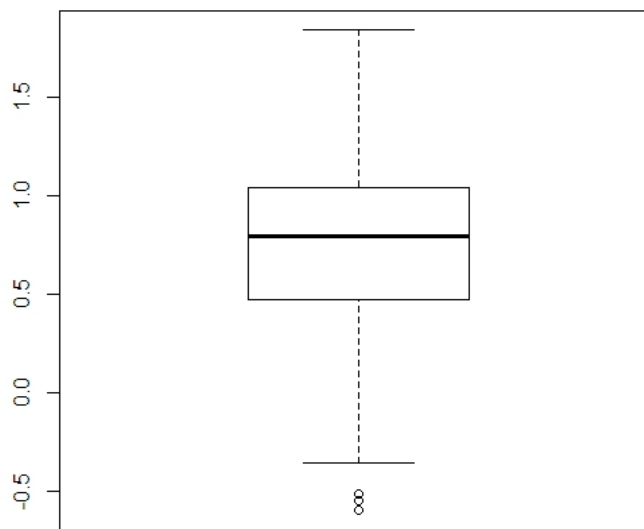


Figura 4.4: gráfico de caja

Al observar la figura 4.4 se puede apreciar fácilmente que hubo un cambio con respecto al gráfico de caja inicial además de un cambio en la escala, pues se ajustó a una más baja. También como se pudo verificar la mediana ubicada en el percentil 50 tiene un valor de 0.79 metros, la máxima altura se posicionó 1.84 metros, la mínima con un valor de  $-0.59$  metros de altura, en el percentil 75 la altura es de 1.04 metros, para el percentil 25 es de 0.47 metros de altura y como se aclaró en el apartado anterior que hacía referencia a estos mismos datos

sin ser tratados con un filtro, la media no se apreciaba fácilmente por lo que se debía calcular, y mostró un valor de 0.74 metros de altura.

Luego, se hace necesario en este punto del análisis mostrar nuevamente el histograma de frecuencia, pues si los estadísticos anteriores pasaron por un cambio en sus valores lógicamente la gráfica asociada a su distribución también cambio de forma parcialmente.

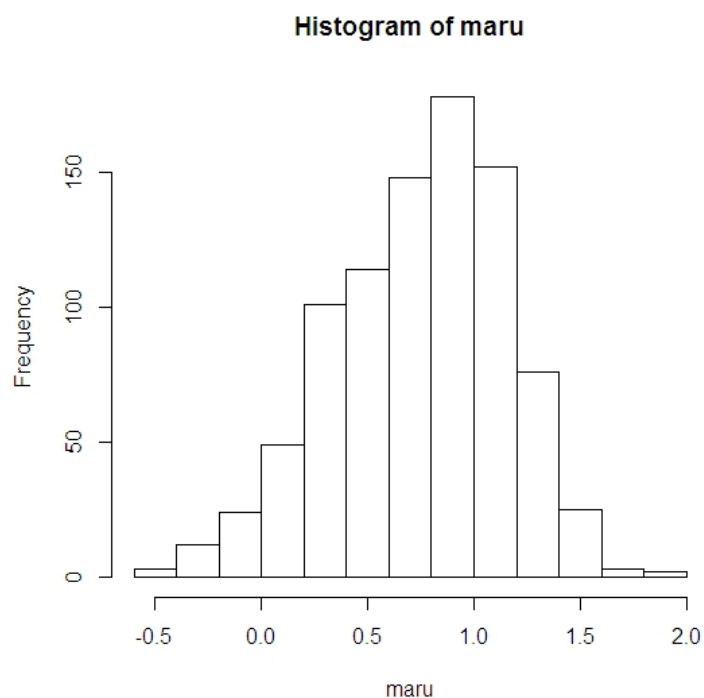


Figura 4.5: Histograma

Este gráfico de caja nos permitió observar cierta simetría en los datos y además al calcular los valores de la media y la varianza se obtuvo que la media tenía un valor de  $\mu = 0.74$  y la varianza un valor de  $\sigma^2 = 0.16$ .

Retomando la serie temporal una vez aplicado el logaritmo a los datos se obtiene la siguiente figura:

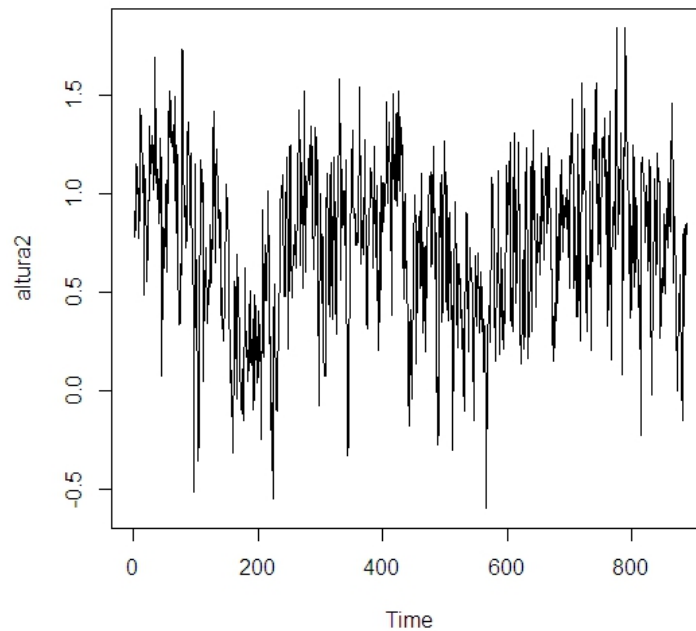


Figura 4.6: Serie Temporal

Al analizar la figura 4.6 (serie temporal filtrada) se aprecia que la misma se suavizó permitiendo localizar algunos patrones de recurrencia o períodos (uno que se aprecia en el gráfico superior y que se muestra mediante las caídas grandes de la serie, que ocurren cada 3 meses y un segundo de cada 16 días que se explicará en detalle más adelante), además en cuanto a su escala también ocurrieron cambios evidentes, es decir, una vez que a los datos se le aplicó la función logaritmo las alturas disminuyeron su tamaño lo cual permitió una mejor visualización de los datos.

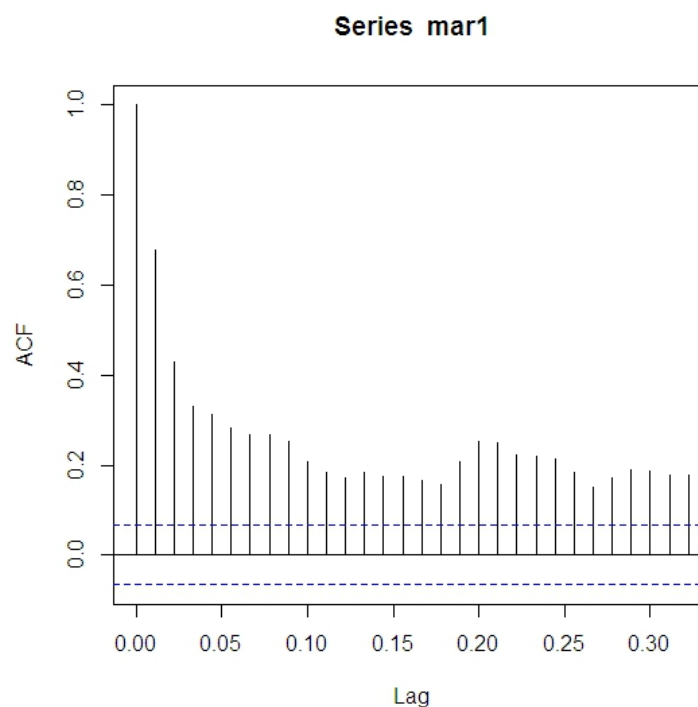


Figura 4.7: función de autocorrelación

Como se muestra en la Figura figura 4.7, que es la función de autocorrelación de la serie temporal de los datos de máximas alturas, existe un segundo patrón de comportamiento con respecto a las alturas y que ocurre cada 16 días, es decir, si contamos las líneas verticales en la función de autocorrelación ocurre que luego de la línea vertical número 16 hay un cambio de comportamiento en las mismas, pues dejan de acortarse sus alturas e inicia la línea 17 con una altura superior a la observada. Esto último nos permitió corroborar la existencia del segundo comportamiento cíclico en los datos.

Luego si hacemos una descomposición aditiva de dicha serie temporal resulta el siguiente conjunto de gráficos:

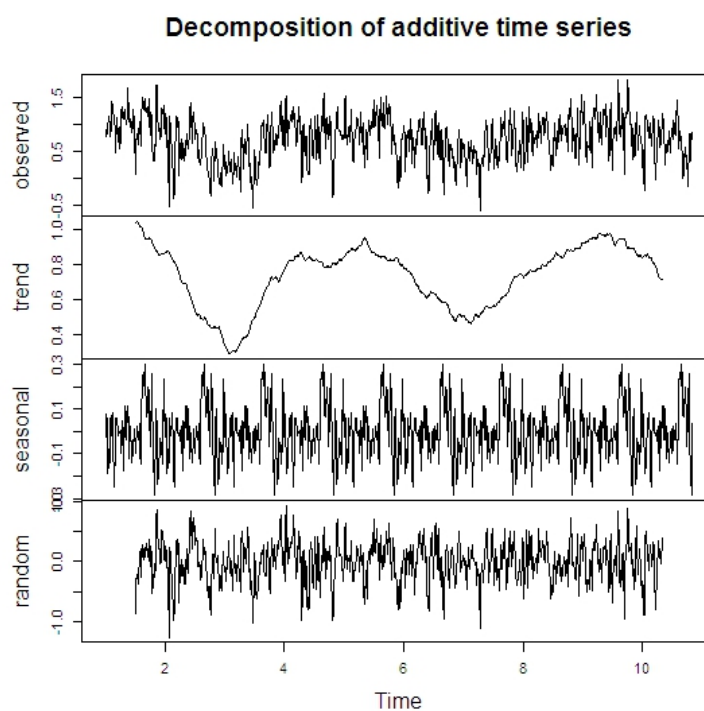


Figura 4.8: Descomposición Aditiva

Notemos que la figura 4.8 nos muestra una descomposición aditiva de la serie temporal (mar) donde en la primera parte del cuadro se muestra la gráfica de la serie temporal, en el segundo cuadro la gráfica de la tendencia de la serie la cual como podemos apreciar no es lineal, para el tercer cuadro está la componente estacional periódico de las máximas alturas el cual se presenta cada 3 meses y por último tenemos el residuo o componente residual de la serie que se obtiene al eliminar la serie original, el patrón periódico y la tendencia antes mencionada.

En este punto planteamos calcular la función de autocorrelación de los residuos de la serie ya que la misma nos ayuda en la comprensión del comportamiento de los datos además de indicarnos si se hace necesario la aplicación de un modelo estadístico que suavizara di-

cho comportamiento y a partir de allí concluir si el residuo de la serie es un ruido blanco gaussiano que no encierra ninguna correlación.

Vista la figura 4.9 se puede afirmar que aún los datos de la serie temporal están correlacionados. Las líneas azules punteadas muestran las cotas entre las cuales deben estar las líneas verticales de color negro, es decir, un indicador de que existe poca correlación o ningún tipo de correlación es que las líneas verticales se encuentren dentro de las cotas azules lo cual en nuestro caso no ocurre por tanto es necesario la implementación de un modelo estadístico que arroje un mejor ajuste de dichos datos.

También se puede apreciar que aún está presente un componente periódico y el mismo está vinculado con el patrón que ya describimos anteriormente de máximos oleajes cada 16 días.

Ahora bien conocida la información anterior procedimos a la implementación de diversos modelos  $ARIMA(p,d,q)$ , los cuales como ya sabemos del capítulo anterior son modelos autoregresivo integrado de media móvil, a la componente residual (random).

Entre los modelos aplicados y con el cual se observó un mejor comportamiento en los residuos fue con el modelo ARIMA de parámetros  $p = 2, d = 0, q = 5$ , es decir, un modelo con componente autoregresiva de orden 2, componente integrada de orden 0 y media móvil de orden 5; sin embargo al aplicar dicho modelo se obtuvieron unos residuos que aun describen un comportamiento periódico (esto se puede apreciar en la primera grafica de la figura 4.10); periodicidad que también es apreciada en la función de autocorrelación de los mismos y además, en el gráfico de esta función también podemos apreciar que existe más de un valor que no está acotado, estos se encuentran en la posición 1 y 17 del eje de las abscisas.

Se incorpora a esta descripción la función de autocorrelación parcial mediante un grá-

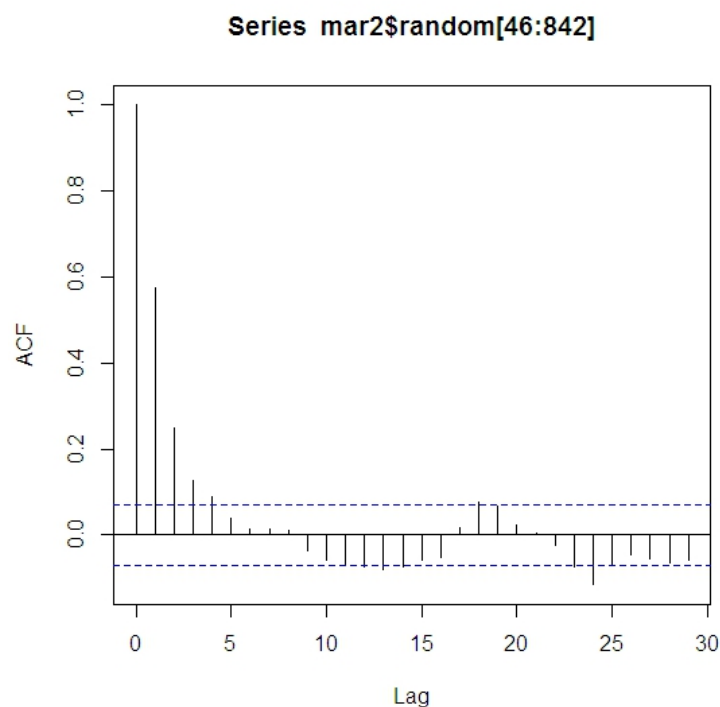


Figura 4.9: función de autocorrelación

fico, pues forma parte de los indicadores del modelo aplicado sobre los residuos y por ende aporta información adicional sobre el mismo.

En la figura 4.11 observamos como existe un comportamiento sucesivo en el patron que lo describe, además de notar que no se encuentra totalmente acotada, pues al menos un dato distinto del primero sobrepasa la línea punteada azul la cual es indicadora de una cota existente, es decir, existe una correlación entre los datos del residuo. Esto último nos hizo proponer un modelo más complejo que para nuestro caso será el modelo **sarima**.

Como se dijo previamente en este punto de la investigación y debido a las característi-

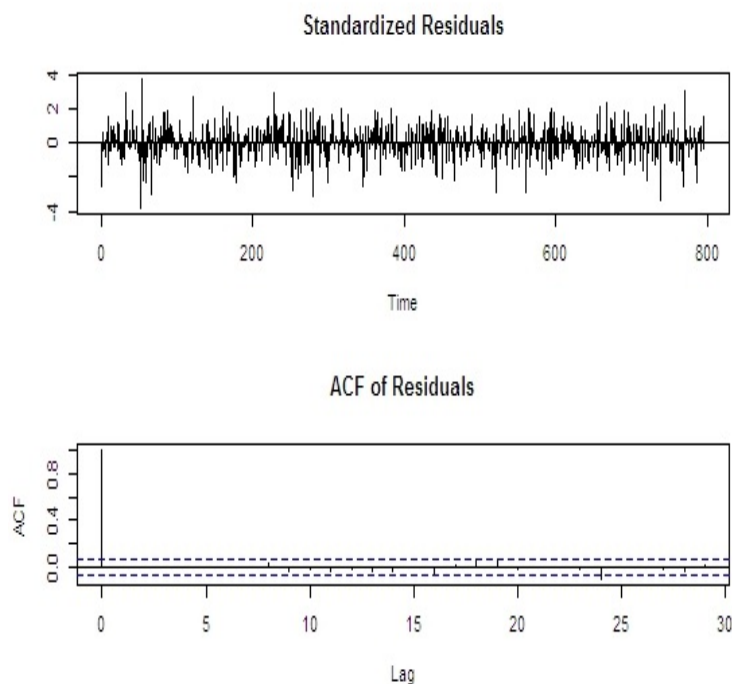


Figura 4.10: Modelo ARIMA

cas que nos muestran las herramientas implementadas anteriormente sobre los datos residuales, hicimos uso del modelo **SARIMA(p,d,q)x(P,D,Q)** que significa modelo autoregresivo de promedio móvil estacional cuyos parámetros de entrada son  $p = 2$ ,  $d = 1$ ,  $q = 6$ ,  $P = 3$ ,  $D = 2$ ,  $Q = 16$  donde **p** es la componente autoregresiva ordinaria de orden 2, **d** componente integrada ordinaria de orden 1, **q** la media móvil ordinaria de orden 6, **P** componente autoregresiva estacional de orden 3, **D** la componente integrada estacional de orden 2 y por último **Q** que es el promedio móvil estacional de orden 2. Este modelo como ya se mencionó en párrafos anteriores se aplicó al residuo de la descomposición aditiva de nuestra serie temporal donde se obtuvieron resultados idóneos:

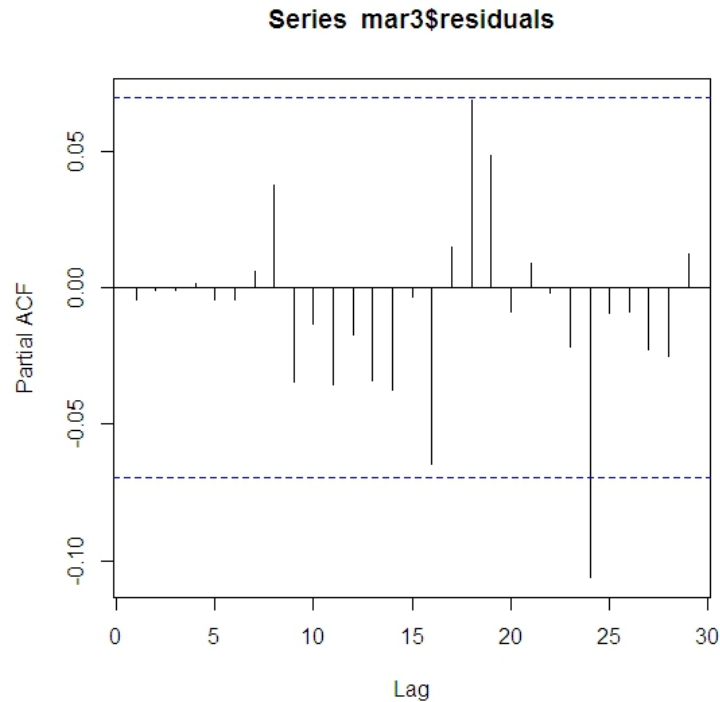


Figura 4.11: función de autocorrelación parcial

Como se puede apreciar en este gráfico los resultados obtenidos con respecto al modelo aplicado sobre el residuo son más congruentes con lo esperado: el primer gráfico muestra la serie temporal del residuo que como se observa no presenta patrones periódicos que nos lleven a concluir sobre la existencia de la correlación entre dichos datos, el segundo indicador es la función de autocorrelación de dichos residuos, el cual muestra un claro acotamiento de los lag (líneas verticales de color negro) con respecto a los valores que muestran las líneas punteadas de color azul. Evento que corrobora la poca correlación entre los datos residuales; el tercer gráfico que se refiere a la "Normal Q-Q plot Residual" muestra como el contraste cuartil de la gaussiana vs cuartil y los datos residuales están sobre la línea recta indicando la gaussianidad de los datos, es decir, que este último modelo planteado es el que

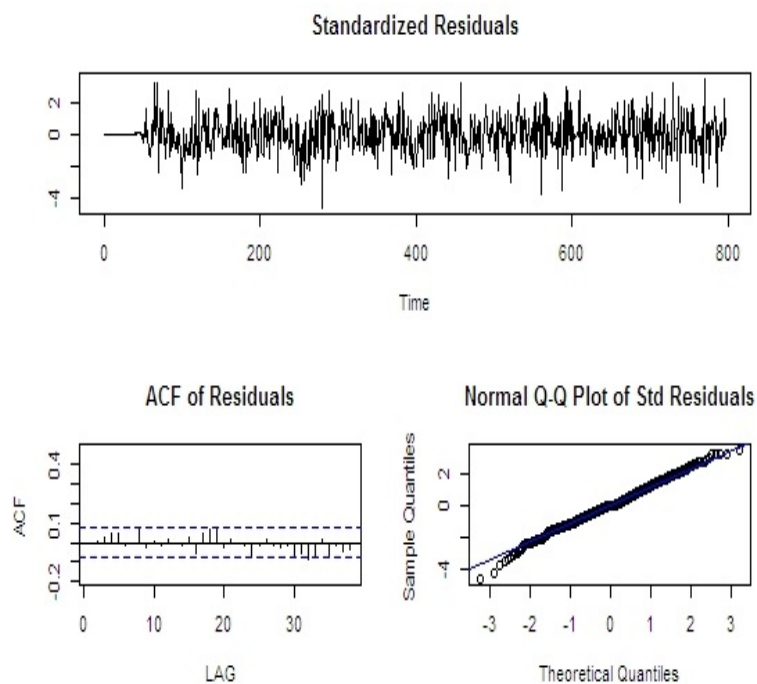


Figura 4.12: Modelo SARIMA

mejor ajusta los datos.

A continuación se presenta un gráfico adicional de la autofunción de correlación parcial de los residuos como parte de los indicadores dentro del modelo planteado que afirman el mejor ajuste de dichos datos.

Notese que éste gráfico nos muestra como los lag se encuentran acotados entre las líneas punteadas azules y además no existe ningún patron evidente en los datos lo que corrobora la no correlación de los datos residuales. Por tanto se hace pertinente la aplicación del modelo aquí propuesto para el ajuste y análisis de nuestra serie temporal (mar).

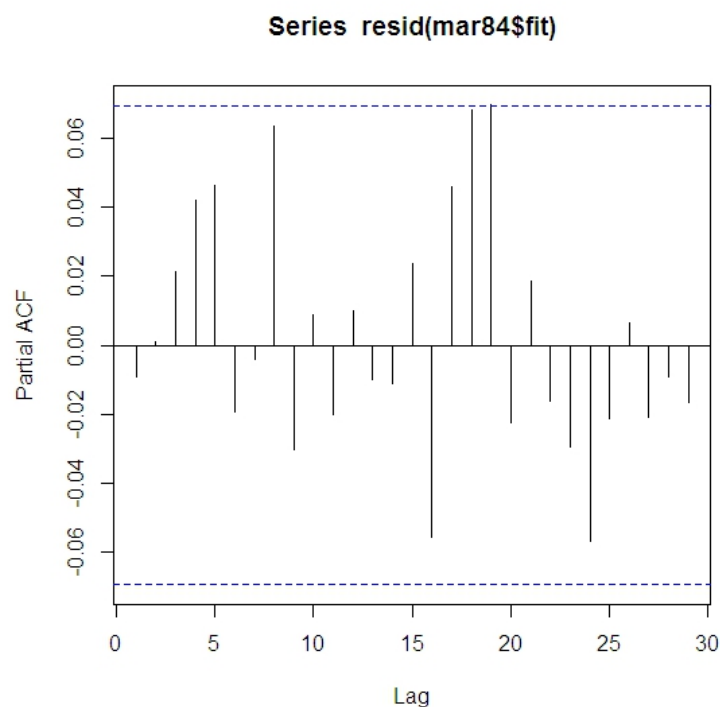


Figura 4.13: función de autocorrelación parcial

En este punto de la investigación se expone un gráfico, que sirve de ayuda en las conclusiones del pronóstico, en el cual se muestra la predicción de los últimos cien (100) días de la data estudiada, aplicando el modelo de serie temporal propuesto, y esta predicción es comparada con los datos originales.

En la figura 4.14 se puede apreciar en negro la serie temporal original de los datos originales y hacia el final una gráfica de color rojo que corresponde a la simulación. Se puede apreciar como la simulación se ajusta bastante bien a los datos originales replicando las características fundamentales de los ciclos de mayor y menor oleaje cada 16 días, así como la tendencia.

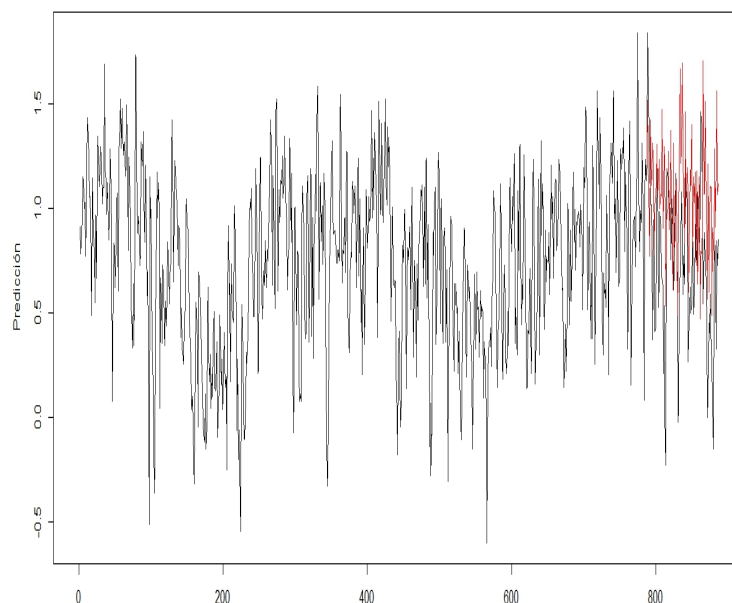


Figura 4.14: predicción 100 días

## Recomendaciones y Conclusiones

Algunas conclusiones y recomendaciones son las siguientes:

- En la costa central de California específicamente en Cabrillo Beach se alcanzan oleajes con alturas máximas los días miércoles con un promedio de 3 metros de altura por día manteniendo el crecimiento las primeras horas del día jueves con un promedio de 3.9 metros de altura máxima. Además presenta fuertes caídas en el oleaje ese mismo día jueves, pero en horas de la noche con un promedio de 0.9 metros.
- Los días viernes se observa un remonte en las alturas del oleaje que se mantiene hasta los días sábados en promedio con alturas máximas de 3.05 metros.

- Las características resaltadas en los ítem 1 y 2 se repiten de forma periódica cada 16 días en ciclos de 3 meses.
- Los ciclos trimestrales de mayor oleaje inician cada año en la última semana de Abril y finalizan la última semana de Julio.

Estos datos pueden ser de gran utilidad en una diversidad de actividades costeras como por ejemplo:

El surf, la pesca de orilla, bodyboard, el kayak, etc.

También se hace de gran importancia resaltar que esta predicción nos ayuda a corroborar el supuesto del cual se partió al momento de iniciar el análisis de los datos como lo fue la hipótesis de gaussianidad con respecto a los mismos.

Entre algunas de las recomendaciones que se pueden hacer con respecto a los resultados obtenidos al realizar el análisis de la serie temporal se encuentran:

- La actividad del mar observada durante el análisis de los datos y su predicción con respecto a los días de mayor altura de olas es una condición suficiente para la práctica de algunos deportes acuáticos.
- Finalmente consideramos que el modelo propuesto si bien ajusta bastante bien a los datos reales tal y como se aprecia en el gráfico de la predicción, el mismo puede ser mejorado utilizando un mayor número de datos.
- La actividad del mar con respecto a grandes olas también arrastra grandes bancos de peces que permiten a los aficionados a la pesca y la industria pesquera obtener variedades que provienen de altamar.

- Esta actividad también advierte a los nadadores y bañistas de esta costa a guardar prudencia al momento de entrar al mar, pues indica el alto riesgo de la zona con respecto a la permanencia en dicho medio.

# Bibliografía

- [1] Brockwell, P.J., and Davis, R.A., (1996) *Introduction to Time Series and Forecasting* SpringerVerlang, New York Inc, New York, pp. 1 – 29, 77 – 110, 81 – 106.
- [2] Brockwell, P.J., and Davis, R.A., (1996) *Introduction to Time Series and Forecasting* SpringerVerlang, New York Inc, New York, pp. 167 – 172, 206 – 213.
- [3] Brockwell, P.J., and Davis, R.A., (1996) *Introduction to Time Series and Forecasting* SpringerVerlang, New York Inc, New York, pp. 301 – 306, 320 – 326.
- [4] Mikosh, T., (1998) *Elementary Stochastic Calculus* Departament of Matematics, University of Groningen, The Netherlands, pp. 14, 23.
- [5] Shumway, R.H and Stoffer, D.S. (2006) *Time Series Analysis and Its Applications with R examples. 2nd. Edition.* Springer. pp. 151 – 159.
- [6] R Core Team (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.