



ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

MODELAMIENTO DE MAREJADAS EN ECUADOR

MODELAMIENTO TEMPORAL DE MAREJADAS EN BAHÍA DE CARÁQUEZ

PROYECTO DE INVESTIGACIÓN

**TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO
REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE INGENIERO
MATEMATICO**

HÉCTOR GASTÓN GUACAPIÑA LLUMIGUSÍN

hector.guacapina@epn.edu.ec

DIRECTOR: YANDIRA DENISSE CUVERO CALERO

yandira.cuvero@epn.edu.ec

DMQ, AGOSTO 2023

CERTIFICACIONES

Yo, HÉCTOR GASTÓN GUACAPIÑA LLUMIGUSÍN, declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

HÉCTOR GASTÓN GUACAPIÑA LLUMIGUSÍN

Certifico que el presente trabajo de integración curricular fue desarrollado por HÉCTOR GASTÓN GUACAPIÑA LLUMIGUSÍN, bajo mi supervisión.

YANDIRA DENISSE CUVERO CALERO
DIRECTOR

DECLARACIÓN DE AUTORÍA

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como los productos resultantes del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

HÉCTOR GASTÓN GUACAPIÑA LLUMIGUSÍN

YANDIRA DENISSE CUVERO CALERO

RESUMEN

Se estudia la relación temporal de la altura de las olas en el puerto de Bahía de Caráquez durante el año 2021. Para lo cual tendremos dos enfoques. El primero, modelar la alturas de las olas como una serie temporal. Un modelo GARCH para la volatilidad de la altura de las olas durante el período de estudio. El segundo, se empleará la escala Douglas, para definir los estados en los cuales puede encontrarse la altura de las olas, con el fin de obtener las probabilidades de transición. A largo plazo se busca obtener la distribución estacionaria de la altura de las olas, con el objetivo de generar un modelo que permita predecir futuros oleajes en favor de la comunidad.

Palabras clave: marejadas, cadena de markov, series de tiempo.

ABSTRACT

The temporal relationship of wave heights in the port of Bahia de Caraquez during the year 2021 is studied. For which we will give two approaches; First, to model the wave heights as a time series. A GARCH model for volatility of the wave heights during the study period. Second, the Douglas scale will be used to define the states in which the waves height can be found, in order to obtain the transition probabilities. In the long term, the aim is to obtain the stationary distribution of waves height, with the objective of generating a model that allows predicting future waves in favor of the community.

Keywords: waves, markov chain, time series.

AGRADECIMIENTO

A mis padres y hermana, que a pesar de las circunstancias siempre han sido mi mayor apoyo y mi motor para llegar a cumplir una meta más.

A Yandira Cuvero por la guía, paciencia y comprensión en la realización del proyecto.

A los profesores de la Facultad de Cienas por los conocimientos impartidos tanto en la parte académica, como moral y profesional; y al personal administrativo que en supieron darnos la ayuda necesaria cuando se necesitaba.

A mis amigos, que compartimos las aulas de clases, compartimos buenos y malos momentos; que sea con una broma o un consejo supimos darnos apoyo para lograr nuestro objetivo.

Y por último y no menos importante a todos los amigos y familiares que no podrán acompañarme en esta meta lograda, pero siempre los recordaré y los tendré presente.

DEDICATORIA

A mis padres, por todo lo que me han dado y la vida les recompensará
y para un buen amigo que fue un gran apoyo en su momento.

Índice general

1. Descripción del componente desarrollado	1
1.1. Objetivo general	2
1.2. Objetivos específicos	2
1.3. Alcance	2
1.4. Marco teórico	2
1.4.1. Serie de tiempo estacionaria	3
1.4.2. Modelos ARIMA	4
1.4.3. Pronóstico	4
1.5. Procesos estocásticos	4
1.5.1. Procesos estacionarios	6
1.5.2. Proceso de Ruido Blanco	7
1.5.3. Estimación de los momentos de procesos estacionarios	8
1.5.4. Cadenas de Markov	10
1.5.5. Escala Douglas	12
2. Metodología	13
2.1. Regón de estudio	13
2.2. INOCAR	14
2.3. Descripción de la base de datos	14
2.4. Análisis Exploratorio	16

2.5. Modelización	17
2.5.1. Serie de tiempo	17
2.5.2. Análisis Residual	19
2.5.3. Validación Cruzada	22
2.5.4. Modelo GARCH	22
2.6. Cadenas de Markov	25
2.6.1. Enero	25
2.6.2. Probabilidad de Tansición $P_{i,j}$	26
2.6.3. Año 2021	30
3. Resultados, conclusiones y recomendaciones	33
3.1. Resultados y Conclusiones	33
3.1.1. Serie de tiempo	33
3.1.2. Cadenas de Makov	35
3.2. Recomendaciones	36
Bibliografía	37

Índice de figuras

2.1. Mapa territorial de Bahía de Caráquez	13
2.2. Diagrama de caja por mes	17
2.3. Altura de marejadas en el año 2021, Bahía de Caraquez	18
2.4. Correlograma y Correlograma Parcial de marejadas año 2021	19
2.5. Análisis residual $ARMA(3,3)$	20
2.6. Análisis residual $ARMA(5,5)$	20
2.7. Análisis residual $ARMA(2,2)$	21
2.8. Residuales al cuadrado del modelo $ARMA(3,3)$	23
2.9. Autocorrelación de Errores al cuadrado	24
2.10 Autocorrelación parcial de Errores al cuadrado	25
2.11 Varianza del modelo $ARMA(3,3)$	26
2.12 Residuales al cuadrado y Varianza del modelo $ARMA(3,3)$	27
2.13 Alturas del mes de Enero	28
2.14 Cadena de markov de alturas del mes de enero	29
3.1. Residuales y varianza del modelo $ARMA(3,3)$	34
3.2. Ajuste y predicción del oleaje mediante el modelo $GARCH(1,1)$ basado en un $ARMA(3,3)$	35

Capítulo 1

Descripción del componente desarrollado

La marejada es un fenómeno caracterizado por el aumento significativo en la altura de las olas, causado por la interacción de diferentes factores [2]. Entre los factores mas influyentes estan:

- Viento: Dado que el viento varia en intensidad y dirección en diferentes momentos y ubicaciones las olas generadas por el viento pueden ser variables y aparentemente aleatorias.
- Marea: El cambio de marea afecta en la profundidad de la ola, lo que lleva a variaciones en la altura de la ola y en su comportamiento diferentes momentos.
- Topografía submarina: la interacción con la topografía submarina, como bancos de arena, arrecifes, pueden generar cambios en la dirección y la altura de las olas, contrinuyendo a la aleatoriedad de la ola.

Por la complejidad de estos factores y su interacción dinámica, las marejadas se modelan de manera más efectiva mediante procesos estocásticos usando tiempos de parada y propiedades de Markov [11].

1.1. Objetivo general

Generar un modelo que permite predecir futuros oleajes en Bahía de Caráquez en 2023.

1.2. Objetivos específicos

1. Comprender la aplicación de cadenas de markov en el modelamiento de la altura de las olas.
2. Realizar el modelamiento temporal de la altura de las olas.
3. Identificar el mejor modelo para modelar olas.

1.3. Alcance

El estudio se restringe al cantón Bahía de Caráquez durante 2020 a 2021.

1.4. Marco teórico

Definición.- Una serie temporal es una sucesión de n observaciones o datos ordenados sobre una característica o varias características de un evento observable. Representación:

$$x_1, x_2, \dots, x_n \quad \text{o} \quad (x_t)_{t=1}^n$$

donde x_t es la observación t -ésima de la serie y n es el número de observaciones que consta la serie completa.

Con el fin de describir la evolución observada, prever la evolución futura de la serie y contrastar sobre alguna de las características o variables a las que esta relacionada dicha serie [10].

La función de autocovarianza mide la dependencia lineal entre dos puntos de una serie. Si $\psi(s, t) = 0$, x_s y x_t no esta relacionadas linealmente.

mente. Es claro que para $s = t$ la autocovarianza se reduce a la varianza, pues:

$$\psi(t, t)E[(x_t - \mu_t)^2] = Var(x_t)$$

Los fenomenos dinámicos que observamos mediante series temporales son: primero los que toman valores estables en el tiempo alrededor de un nivel constane, sin mostrar una tendencia a largo plazo. Por ejemplo cantidad de lluvia mensuales o anuales en una determinada región, la temperatura maxima o media durante el transcurso de un año, la altura de las olas en una determina región durante un periodo de estudio, entre otros, a estos procesos se denominan estacionarios y segundo series no estacionarias que presentan tendencia, estacionalidad, entre otros efectos en el transcurso del tiempo. Notemos que una serie puede ser estable o estacionario en un período de tiempo corto y no estacionaria a largo plazo, es decir que entre más amplio el período de estudio, es menos probable que la serie sea estable [4].

1.4.1. Serie de tiempo estacionaria

Definción.- Sea $(x_t)_{t=1}^n$ una serie de tiempo es estrictamente estacionaria si la distribución de toda coleccion de valores

$$\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\}$$

es identica al conjunto desplazado en el tiempo

$$\{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}$$

es decir,

$$P(x_{t_1} \leq c_1 \dots, x_{t_k} \leq c_k) = P(x_{t_1+h} \leq c_1 \dots, x_{t_k+h} \leq c_k)$$

para todo $k = 1, 2, \dots$, en los tiempos t_1, t_2, \dots, t_k , todos los números c_1, c_2, \dots, c_k y todo los desplazo de tiempo $h \in \mathbb{N}$ [10].

1.4.2. Modelos ARIMA

La regresión clásica suele ser insuficiente para explicar todas las dinámicas interesantes de una serie temporal. Por ejemplo, el ACF de los residuos de la simulación revela una estructura adicional en los datos que la regresión no capturó. En cambio, la introducción de la correlación como un fenómeno que puede generarse a través de relaciones lineales rezagadas lleva a proponer los modelos autorregresivo (AR) y autorregresivo de promedio móvil (ARMA) [1]. Agregar modelos no estacionarios a la mezcla conduce al modelo de promedio móvil integrado autorregresivo (ARIMA) popularizado en el trabajo histórico de Box y Jenkins [10]. El método de Box Jenkins para identificar un modelo ARIMA mediante técnicas para la estimación de parámetros y el pronóstico para estos modelos.

1.4.3. Pronóstico

En el pronóstico, el objetivo es predecir los valores futuros de una serie de tiempo, x_{n+m} , $m = 1, 2, \dots$, en base a los datos recopilados hasta el presente, $x = \{x_n, x_{n-1}, \dots, x_1\}$

El predictor de error cuadrático medio mínimo de x_{n+m} es

$$x_{n+m}^n = E(x_{n+m} | \mathbf{x})$$

porque la expectativa condicional minimiza el error cuadrático medio [8].

1.5. Procesos estocásticos

Definición.- Un proceso estocástico es una colección de variables aleatorias, $\{Y_t : t \in T\}$ ordenadas y equidistantes cronológicamente, referidas a una o varias características observables en diferentes momentos. Donde T es el espacio paramétrico.

Representación:

$$\dots, Y_{-1}, Y_0, Y_1, Y_2 \dots$$

Donde Y_t es una variable aleatoria escalar en el momento t referida a la

característica observable. Mediante una abstracción se tendría:

$$\begin{aligned} \text{Proceso estocástico: } & \dots, Y_{-1}, Y_0, \underbrace{Y_1, Y_2, \dots, Y_n}_{}, Y_{n+1}, \dots \\ \text{Serie temporal: } & y_1, y_2, \dots, y_n \end{aligned}$$

Tomando el espacio paramétrico como $T = \{0, 1, 2, \dots\}$, que notarían tiempos. Un proceso estocástico a tiempo discreto se denotará por $\{Y_n : n = 0, 1, \dots\}$ el estado del sistema al tiempo n .

Un proceso a tiempo continuo se denotará por $\{Y_t : t \geq 0\}$

Notemos que los estados considerados son subconjuntos \mathbb{Z} y más general el conjunto de estados de números reales \mathbb{R} [3].

Entre otros procesos a mencionar se tienen los siguientes:

■ **Procesos de Markov**

Los estados anteriores no tienen influencia en los estados futuros del sistema. Se puede expresar como:

$$y_0, y_1, \dots, y_{n-1}(\text{pasado}), y_n(\text{presente}), y_{n+1}(\text{futuro})$$

cumple:

$$P(Y_{n+1} = y_{n+1} | Y_0 = y_0, \dots, Y_n = y_n) = P(Y_{n+1} = y_{n+1} | Y_n = y_n)$$

Es decir que el estado o evento ($Y_{n+1} = y_{n+1}$) futuro, depende únicamente del evento ($Y_n = y_n$), entonces la información de los eventos pasados ($Y_0 = y_0, \dots, Y_n = y_n$) es irrelevante. A este proceso se denomina *proceso de Markov*[8].

■ **Procesos con incremento independientes**

Un proceso estocástico a tiempo continuo $\{Y_t : t \geq 0\}$, tiene incrementos independientes si para cualquier tiempo $0 \leq t_1 < t_2 < \dots < t_n$, las variables $Y_{t_1}, Y_{t_2} - Y_{t_1}, \dots, Y_{t_n} - Y_{t_{n-1}}$ son independientes. Es decir los desplazamientos del proceso en los intervalos de tiempo disjuntos son independientes entre sí.

■ **Procesos estacionarios**

Un proceso estocástico a tiempo continuo $\{Y_t : t \geq 0\}$ es estacionario en sentido estricto si para cualquiera tiempo t_1, \dots, t_n , la distribución del vector $(Y_{t_1}, \dots, Y_{t_n})$ es la misma que el vector $(Y_{t_1+h}, \dots, Y_{t_n+h})$ para cualquier valor $h > 0$ [3].

■ Procesos con incrementos estacionarios

Un proceso estocástico a tiempo continuo $\{Y_t : t \geq 0\}$ tiene incrementos estacionarios si para cualquier tiempo $s < t$ y $h > 0$, las variables $Y_{t+h} - Y_{s+h}$ y $Y_t - Y_s$ presentan la misma distribución.

1.5.1. Procesos estacionarios

La obtención de las distribuciones de probabilidad del proceso es posible en ciertas situaciones, por ejemplo con variables climáticas, bajo el supuesto de cada año se realice una observación del proceso o las técnicas. Sin embargo, situaciones de interés tales como económicas o sociales solo podemos realizar una observación del proceso. Por ejemplo la serie económica de la riqueza de un país, el proceso existe conceptualmente pero no se puede obtener muestras sucesivas o independientes del mismo[7].

Un proceso estocástico es estacionario en sentido estricto si:

1. Las distribuciones marginales de todas las variables son idénticas
2. Las distribuciones de cualquier conjunto de variables solo depende de los retardos entre si.

Estas dos condiciones se resumen estableciendo:

$$F(y_i, y_j, \dots, y_k) = F(y_{i+h}, y_{j+h}, \dots, y_{k+h}) \quad \text{para cualquier valor } h > 0$$

La estacionariedad estricta es una condición muy fuerte, ya que es necesario de conocer las distribuciones conjuntas para la selección de variables del proceso[3]. Una condición más débil y en la práctica más sencillo de contrastar, estacionariedad en sentido débil, que implica la media, varianza y estructura de covarianzas estables a lo largo del tiempo. Es decir:

1. $\mu_t = \mu = cte$
2. $\sigma_t^2 = \sigma^2 = cte$
3. $\gamma(t, t - k) = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k \quad k = 0, \pm 1, \pm 2, \dots$

En un proceso estacionario las autocovarianzas y autocorrelaciones solo depende del retardo entre observaciones[1], entonces:

$$Cov(y_t, y_{t+k}) = Cov(y_{t+j}, y_{t+j+k}) = \gamma_k \quad k = 0, \pm 1, \pm 2, \dots$$

y las autocorrelaciones:

$$\rho_k = \frac{Cov(y_t, y_{t-k})}{\sqrt{var(y_t)var(y_{t-k})}} = \frac{\gamma_k}{\gamma_0}$$

Por tanto, en procesos estacionarios se tiene: $\gamma_0 = \sigma^2$, $\gamma_k = \gamma_{-k}$ y $\rho_k = \rho_{-k}$

1.5.2. Proceso de Ruido Blanco

El proceso de *ruido blanco* está definido por las siguientes condiciones:

1. $E[y_t] = 0, \quad t = 1, 2, \dots$
2. $Var(y_t) = \sigma^2, \quad t = 1, 2, \dots$
3. $Cov(y_t, y_{t-k}) = 0, \quad k = \pm 1, \pm 2, \dots$

En estos procesos conocer los valores pasados no proporciona ninguna información sobre el futuro, pues el proceso *no tiene memoria*[3].

Un proceso estacionario con varianza finita σ^2 y que verifique:

$$E(y_t | y_{t-1}, \dots, y_1) = 0$$

es una equivalencia al ruido blanco, entonces puede demostrarse que se verifica las condiciones 1 y 3 anteriores.

Un proceso de ruido blanco no necesariamente es estacionario en sentido estricto ni puede estar formado por variables independientes, por el

hecho que solo exige incorrelación. Bajo el supuesto de independencia el proceso se denomina *ruido blanco estricto*, además bajo el supuesto de que las variables tienen distribución normal resulta un proceso *ruido blanco normal*[8].

1.5.3. Estimación de los momentos de procesos estacionarios

Un proceso estacionario con media $\mu = E[y_t]$, varianza $\sigma^2 = \gamma_0 = \text{Var}(y_t)$ y covarianzas $\text{Cov}(y_t, y_{t-k}) = \gamma_k$ del cual se observa una realización (y_1, \dots, y_T) [7], estudiaremos como estimar los momentos en base a la observación dada

■ Estimación de la media

Un estimador de la media poblacional es la media muestral (\bar{y}) :

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

se verifica que:

$$E(\bar{y}) = \frac{1}{T} \sum_{t=1}^T E(y_t) = \mu$$

Para datos independientes la varianza de la media muestral como estimador de la media poblacional es σ^2/T por lo tanto al aumentar el tamaño de muestra el error cuadrático medio de la estimación dado por

$$E(\bar{y} - \mu)^2$$

tiende a cero, dado que coincide con la varianza de \bar{y} . Cabe mencionar que un proceso estocástico estacionario la independencia no necesariamente es cierta y puede darse el caso que al aumentar el tamaño de muestra la varianza de la estimación de la media no tienda a cero[2].

■ Ergodicidad

Decimos que un proceso estocástico es *ergódico* para la estimación si verifica:

$$\lim_{T \rightarrow \infty} E(\bar{y} - \mu)^2 \rightarrow 0$$

Como el estimador \bar{y} es centrado para μ , el error cuadrático medio coincide con la varianza del estimador tal que:

$$var(\bar{y}) = E(\bar{y} - \mu)^2 = \frac{1}{T^2} E \left(\sum_{t=1}^T (y_t - \mu) \right)^2$$

Podemos reescribir como:

$$var(\bar{y}) = \frac{1}{T^2} \left[\sum_{t=1}^T E(\bar{y} - \mu) + 2 \sum_{i=1}^T \sum_{j=i+1}^T E((y_i - \mu)(y_j - \mu)) \right]$$

De donde el primer término es $T\sigma^2$, el segundo término contiene un doble sumatorio con $T - 1$ veces las covarianzas de orden 1, $T - 2$ veces las covarianzas de orden 2, en general, $T - i$ veces las covarianzas de orden i . Por lo tanto puede escribirse como

$$var(\bar{y}) = \frac{1}{T} \left[\sigma^2 + 2 \sum_{i=1}^{T-1} \left(1 - \frac{i}{T}\right) \gamma_i \right] \quad (1.1)$$

Por lo tanto si las covarianzas son cero, se tiene un proceso de ruido blanco con varianza de la media muestral σ^2/T , caso contrario la varianza muestral puede ser mayor que para observaciones independientes[3]. La condición para que $var(\bar{y})$ tienda a cero al aumentar T , es que converja a una constante al aumentar el tamaño de muestra T . Una condición necesaria de que el sumatorio converja es:

$$\lim_{i \rightarrow \infty} \gamma_i \rightarrow 0$$

es decir la dependencia entre observaciones tiende a cero al aumentar el retardo.

Aunque no es una condición suficiente, por ejemplo: una serie constante tiene correlación uno entre las observaciones y no es ergódica.

Utilizando autocorrelaciones en la ecuación (1.1) puede escribirse

como:

$$\text{var}(\bar{y}) = \frac{\gamma_0}{T} \left[1 + 2 \sum_{i=1}^{T-1} \left(1 - \frac{i}{T}\right) \rho_i \right]$$

Donde $\rho_i = \gamma_i/\gamma_0$ son las autocorrelaciones del proceso. Similar como con las autocovarianzas, para que el proceso sea ergódico una condición necesaria es que en la estimación de la media $\lim_{i \rightarrow \infty} \rho_i \rightarrow 0$ [3].

■ Estimación de las autocovarianzas y autocorrelaciones

Suponiendo que la media es conocida del proceso[1], el estimador de autocovarianza de orden k es:

$$\hat{\gamma}_k = \frac{1}{T-k} \sum_{t=k+1}^T (y_t - \mu)(y_{t-k} - \mu)$$

1.5.4. Cadenas de Markov

Las cadenas de Markov, es un modelo suficientemente complejo para describir ciertas características no triviales y lo suficientemente sencillo para ser analizado matemáticamente[3], además de la amplia gama de fenómenos científicos y sociales a la cual pueden aplicarse.

■ Propiedad de Markov

Notaremos la probabilidad $P(Y_n = y_n)$ como $p(y_n)$, donde le subíndice hace referencia a la variable, análogamente con la probabilidad condicional $P(y_{n+1}|y_n)$, para una mejor descripción de la propiedad.

Definición.- Una cadena de Markov es un proceso estocástico a tiempo discreto $\{Y_n : n = 0, 1, \dots\}$ y satisface la propiedad de Markov, tal que para cualquier entero $n \geq 0$ y para cualquier estado y_0, \dots, y_{n+1} se cumple que:

$$p(y_{n+1}|y_0, \dots, y_n) = p(y_{n+1}|y_n) \quad (1.2)$$

Establece que la distribución de probabilidad del estado del proceso al

tiempo futuro $n+1$ depende unicamente del estado del proceso al tiempo n y no de los estados en tiempos pasados $0, 1, \dots, n-1$ [8].

■ **Probabilidades de Transición**

Sean i y j dos estados de una cadena de Markov. A la probabilidad

$$P(Y_{n+1} = j | Y_n = i)$$

la notaremos $p_{ij}(n, n + 1)$ representa la probabilidad de transición de ir del estado i en el tiempo n , al estado j en el tiempo $n+1$ [8]. Se denominan probabilidades de transición en un paso. Si no dependen de n entonces la cadena es estacionaria en el tiempo, y la notaremos p_{ij} . Por ejemplo si el conjunto de estados $\{0, 1, 2, \dots\}$ se obtiene la matriz de transición en un paso

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \dots \\ p_{10} & p_{11} & p_{12} & \dots \\ p_{20} & p_{21} & p_{22} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Es decir la entrada (i,j) de esta matriz es la probabilidad de transición p_{ij} de pasar del estado i al estado j en una unidad de tiempo.

■ **Propiedad 1.** La matriz de probabilidades de transición $P = (p_{ij})$ cumple:

- a) $p_{ij} \geq 0$
- b) $\sum_j p_{ij} = 1$

Debido a la propiedad de Markov, dicha matriz captura la esencia del proceso y determina el comportamiento de la cadena en cualquier tiempo futuro[3].

■ **Probabilidades de transición en n pasos**

La probabilidad $P(Y_{n+m} = j | Y_n = i)$ representa la probabilidad de pasar del estado i al tiempo n , al estado j al tiempo $n+m$. Bajo el supuesto de estacionariedad esta probabilidad no depende realmente

[3], por lo cual $P(Y_n = j | Y_0 = i)$ y se denota $p_{ij}(n)$ o $p_{ij}^{(n)}$. La matriz de probabilidades de transición en n pasos es:

$$P(n) = \begin{pmatrix} p_{00}(n) & p_{01}(n) & p_{02}(n) & \dots \\ p_{10}(n) & p_{11}(n) & p_{12}(n) & \dots \\ p_{20}(n) & p_{21}(n) & p_{22}(n) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Si $n = 1$, es la matriz de transición a un paso. Además, cuando $n = 0$ se define $p_{ij}(0)$ como la función delta de Kronecker:

$$p_{ij}(0) = \delta_{ij} = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$$

Es decir, que la cadena después de realizar 0 pasos, debe estar en su estado de partida.

1.5.5. Escala Douglas

La Escala Douglas clasifica los diferentes estados del mar en 9 niveles o grados, contiene las características del oleaje, el tamaño y la denominación de las olas tomando como referencia la altura del oleaje.

Grado	Altura de la ola (m)	Oleaje	Denominación
0	sin olas	sin oleaje	Mar en calma
1	0 a 0.10	Muy bajo	Mar rizada
2	0.10 a 0.5	Bajo	Marejadilla
3	0.5 a 1.25	Ligero	Marejada
4	1.25 a 2.5	Moderado	Fuerte Marejada
5	2.5 a 4	Grueso moderado	Gruesa
6	4 a 6	Grueso	Muy gruesa
7	6 a 9	Alto	Arbolada
8	9 a 14	Muy alto	Montañosa
9	Más de 14	Confuso	Enorme

Cuadro 1.1: Escala Douglas

Los cambios repentinos en las marejadas permiten alertar a los puertos y tomar medidas adecuadas en favor de la comunidad[9].

Capítulo 2

Metodología

2.1. Regón de estudio

Bahía de Caraquez, se localiza en el centro de la región litoral del Ecuador, Provincia de Manabí, a una altitud de 15 msnm y con un clima tropical de 25°C en promedio. La ciudad es el núcleo del area metropolitana, la cual esta constituida por ciudades y parroquias rurales cercanas, albergando más de 40.000 habitantes y una territorial de 181,75 km^2 , se muestra en la figura 2.1.

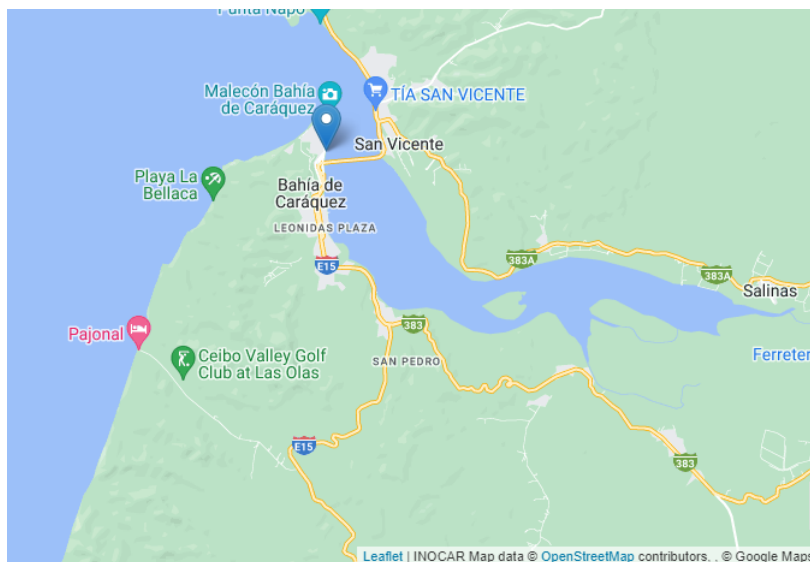


Figura 2.1: Mapa territorial de Bahía de Caraquez

2.2. INOCAR

En febrero de 1932 se crea el Servicio Hidrográfico, con el objetivo de realizar levantamientos hidrográficos, creando las primeras cartas náuticas de la costa ecuatoriana. En julio de 1972, sube a la categoría de Instituto Oceanográfico, con la tarea de proporcionar seguridad a la navegación y compilar la cartografía náutica nacional. El 8 de mayo de 2020, se fusiona con el Instituto Antártico convirtiéndose en el Instituto Oceanográfico y Antártico de la Armada (INOCAR) desde entonces realiza levantamientos hidrográficos con el fin de proporcionar seguridad a la navegación y cumplir la cartografía náutica nacional, además de caracterizar los espacios marítimos de interés e implementar sistemas de señalización, siendo el encargado de realizar el monitoreo, mediciones y predicciones de fenómenos oceanográficos en las costas ecuatorianas [5].

La misión consiste en caracterizar los espacios marítimos de interés nacional e implementar operativos los sistemas de señalización marítima. Los objetivos principales del INOCAR son:

- Incrementar la contribución en favor de la seguridad en la navegación en el espacio marítimo nacional.
- Contribuir y posicionar a Ecuador dentro de la Cuenca del Pacífico y Sistema del Tratado Antártico.

2.3. Descripción de la base de datos

La base de datos, contiene 1410 datos de altura correspondientes al año 2021, conformada por las variables:

- Mes: variable categórica - meses del año
- Día: variable categórica - día del mes
- Hora: variable numérica - hora a la que se toma el dato de altura
- Altura: variable numérica - altura de la ola en metros.

Se tomó el promedio cada 6 horas de la altura, debido a que la medición se lo realiza mediante la boya colocada por el INOCAR. Por lo cual, ciertos días tiene solo 3 datos y 4 datos. Se realiza la codificación de la altura de olas y se guarda en una variable categórica denominada Grado con valores entre 0 y 9 (niveles), en base a la Escala Douglas de marejadas.

El documento pdf con los datos se descargo de la página del INOCAR. Al ser un documento en pdf no tenia la estrucutra necesaria, para lo cual se realizaron los siguientes pasos:

1. Descargar los datos en formato pdf de la tablas de mareas de la página del INOCAR correspondientes al año 2021.
2. Mediante el software en línea, I love PDF [6], se realizó la transformación del documento en pdf a un documento en Excel.
3. Realizar una primera limpieza de los datos obtenidos, enfocándose en los datos atípicos que se generaron del software en línea, para poder estructurar los datos e identificar las variables: la altura de las olas, la hora, el día y el mes.
4. Cargar la base de datos al software R, para realizar una segunda limpieza de datos enfocándose en las características de las variables, es decir datos que presenten un formato diferente al formato de la variable analizada, tenga inconsistencia, además de revisar si se encontraban valores perdidos o NA's.

El procesamiento completo desde el momento que se obtuvo los datos de marejadas del INOCAR hasta obtener la base datos realizada la limpieza de datos en el software R, se lo hizo por etapas:

Descargar los datos en formato pdf de la página del INOCAR, se obtuvo 1 documento en PDF con todos los datos lo cual con llevo 2 días de trabajo, para su posterior etapa de transformación a Excel con llevo 1 día de trabajo y se generó 4 documentos en formato Excel de la transformación, En una semana de trabajo se realizó la primera limpieza de datos atípicos en Excel, se generó 1 documento en formato *xlsx* con la limpieza de los datos y 1 documento adicional

en *xlsx* estructurado la base datos inicial con las variables mencionadas. Por último, se cargó en el software en R, se realizó la última limpieza de valores perdidos o NA's y dando como resultado la base datos final.

2.4. Análisis Exploratorio

Calculando las medidas de tendencia central, valores de mínimo y máximo de las alturas de las olas por mes durante el año 2021, los resultados se muestran el cuadro 2.1. La media de las alturas de todo los meses son similares entre sí y teniendo en cuenta que la altura máxima de los meses del año 2021 fue de 3,2 [m].

Mes	n	Media	Min	Max	Mediana
Enero	119	1.56	0.1	2.9	2.00
Febrero	109	1.53	0.1	3.0	1.10
Marzo	119	1.56	0.0	3.2	2.10
Abril	116	1.56	-0.1	3.2	1.60
Mayo	120	1.57	-0.1	3.1	1.60
Junio	116	1.56	0.0	3.0	1.55
Julio	120	1.54	0.1	2.9	1.55
Agosto	120	1.54	0.1	2.9	1.55
Septiembre	116	1.54	0.1	3.1	1.55
Octubre	119	1.56	0.0	3.1	1.10
Noviembre	116	1.57	-0.1	3.2	1.60
Diciembre	120	1.56	0.0	3.1	1.55

Cuadro 2.1: Estadísticas Descriptivas del año 2021

Ahora en base al diagrama de caja de los meses del año 2021 se observa en la figura 2.2, no presenta datos atípicos, durante los meses enero, febrero, marzo y octubre respectivamente el valor de la mediana de alturas de las olas difieren de los demás meses.

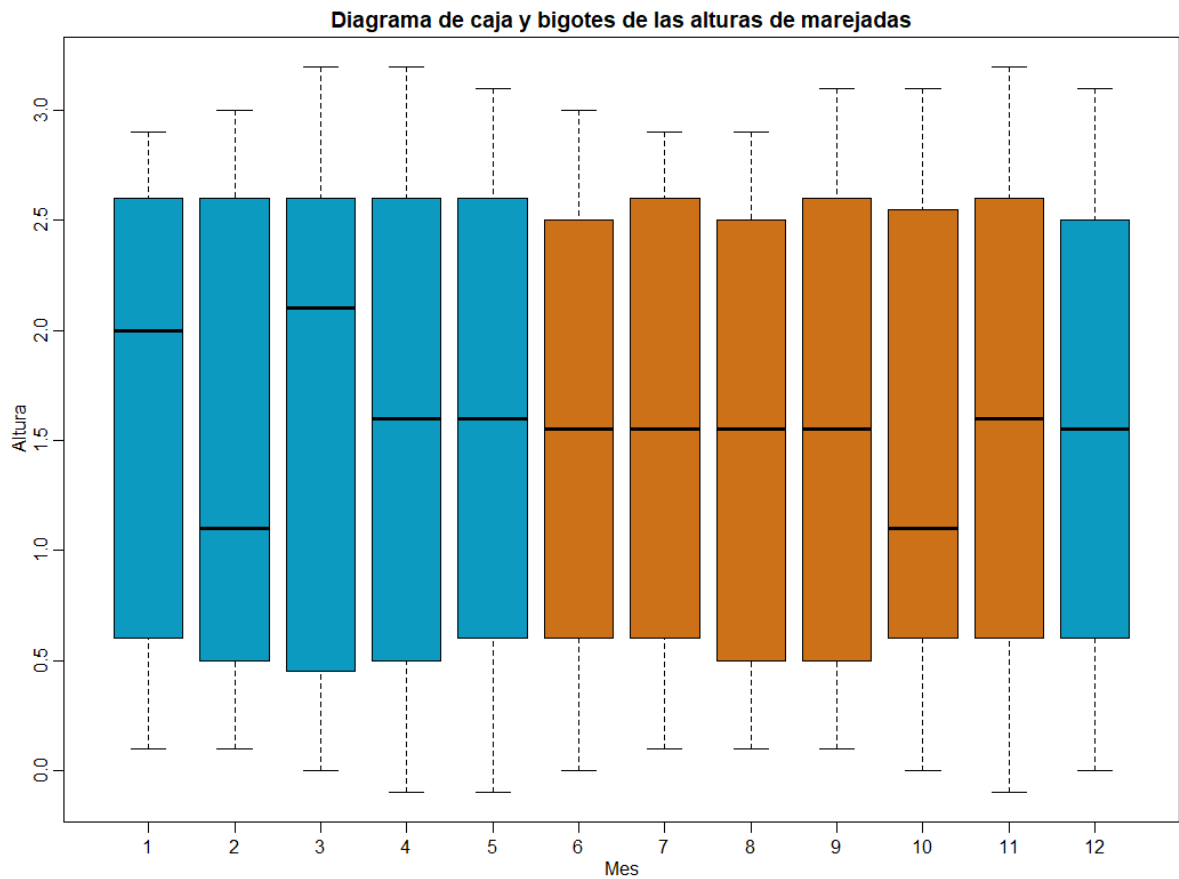


Figura 2.2: Diagrama de caja por mes

2.5. Modelización

Para la modelización de los datos tomaremos un enfoque temporal basado en una serie de tiempo y cadena de markov.

2.5.1. Serie de tiempo

El análisis correspondiente a series temporales, parte de la visualización de los datos que conforman la base de datos, para nuestro caso particular la altura de las olas. A continuación mostraremos la serie temporal de la altura de olas durante el año 2021, se observa en la figura [2.3](#)

Notemos que existe una periodicidad en la serie puesto que ciertos meses del año, se encuentran en una misma estación (invierno o verano) generando alturas similares en las olas.

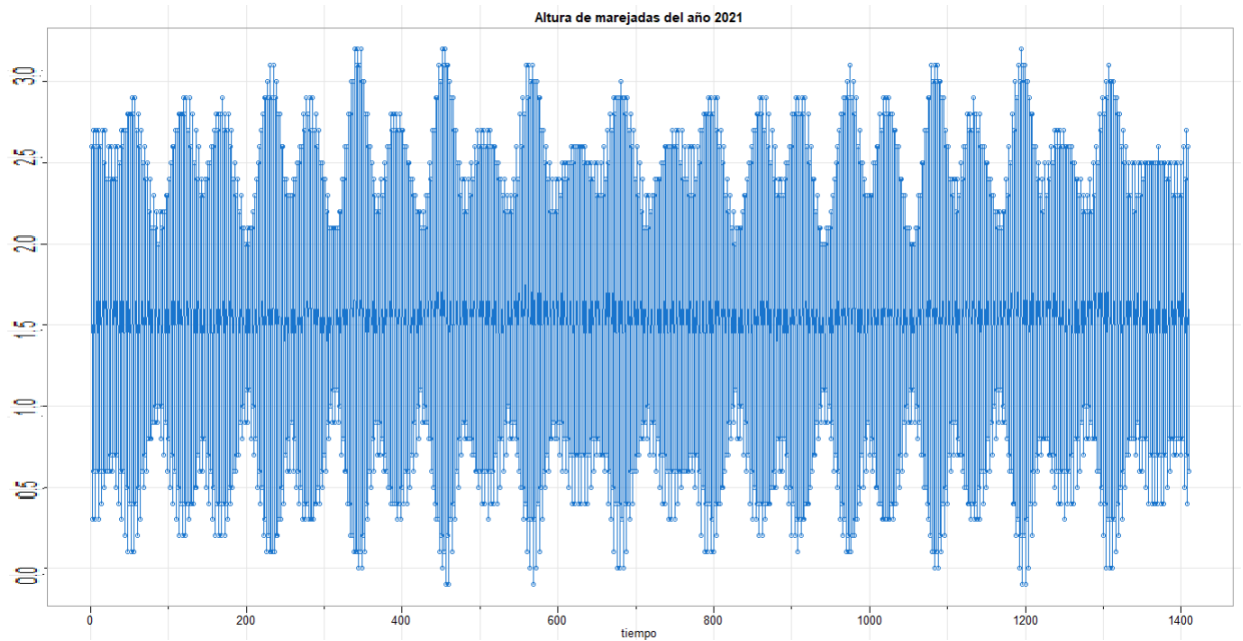


Figura 2.3: Altura de marejadas en el año 2021, Bahía de Caraquez

Mediante el test de Dickey-Fuller, contrastaremos si la serie presenta una raíz unitaria para definir si es una serie estacionaria o en contraste si no lo es, los resultados del test lo obtenemos mediante el software R (anexo ??). En base al criterio del *valor*, y mediante el software R se obtiene $valor = 0,01 < 0,05$ (Nivel de significancia), concluimos que la serie temporal de alturas de las olas no presenta una raíz unitaria, por lo tanto, es una serie estacionaria.

En base al correlograma ACF y correlograma parcial PACF (ver la figura 2.4), nuestros modelos a evaluar serían un ARMA(2,2), ARMA(3,0), ARMA(3,3), ARMA(5,0) y ARMA(5,5).

Mediante el software R, desarrollamos los modelos planteados y en base al Criterio de Akaike (AIC), Sesgo Corregido (AICc) e Información Bayesiano (BIC) resumidos en el cuadro 2.2. Observamos que en base a los tres criterios el mejor modelo es un ARMA(5,5) con $AIC = -3,192$, $AICc = -3,192$ y $BIC = -3,147$ en comparación con los valores de los demás modelos, es menor.

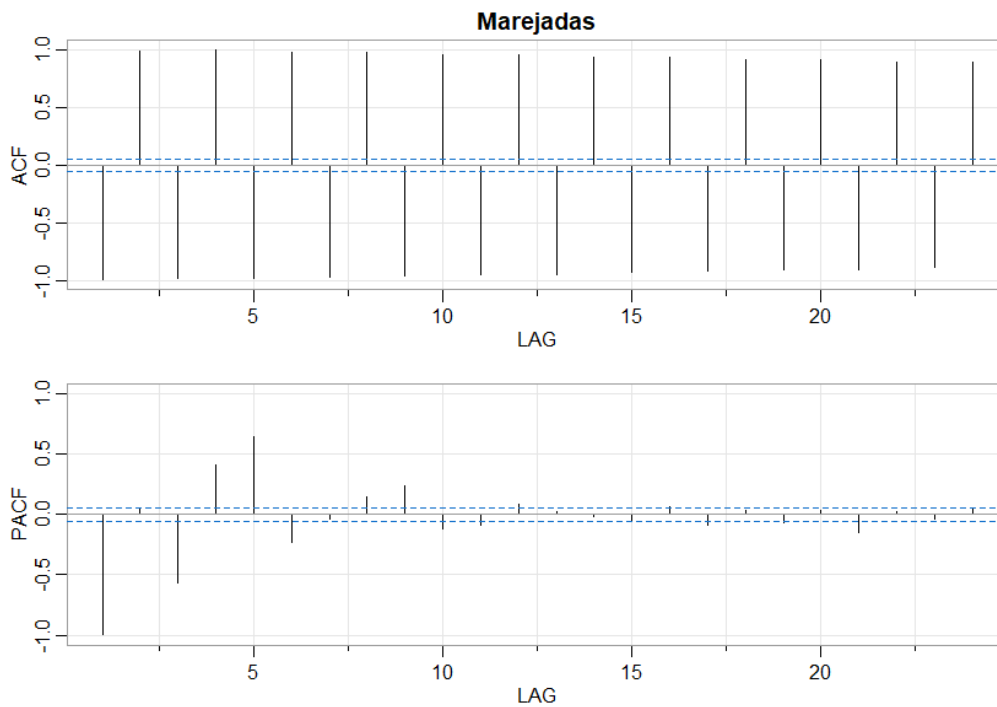


Figura 2.4: Correlograma y Correlograma Parcial de marejadas año 2021

MODELO	AIC	AICc	BIC
ARMA(2,2)	-1.322	-1.322	-1.299
ARMA(3,3)	-2.600	-2.60	-2.571
ARMA(3,0)	-1.787	-1.787	-1.765
ARMA(5,0)	-2.825	-2.858	-2.799
ARMA(5,5)	-3.192	-3.192	-3.147

Cuadro 2.2: Resumen de Modelos ARMA

2.5.2. Análisis Residual

A continuación, presentaremos los residuos para el mejor modelo y un modelo para su comparación, los residuos del modelo ARMA (3,3),

En la figura 2.5, observamos que los residuos estandarizados se encuentran centrado en 0, con el detalle que se encuentra con una gran variación en los mismos, además existe una gran cantidad de rezagos estan por fuera del rango, por lo tanto los residuos no son independientes. Cabe mencionar que los residuos siguen una distribución normal. Ahora realizando el análisis residual del mejor modelo, $ARMA(5,5)$, en base a los tres criterios anteriores se presenta en la figura 2.6

En comparación al modelo ARMA(3,3), los residuos estandarizados tie-

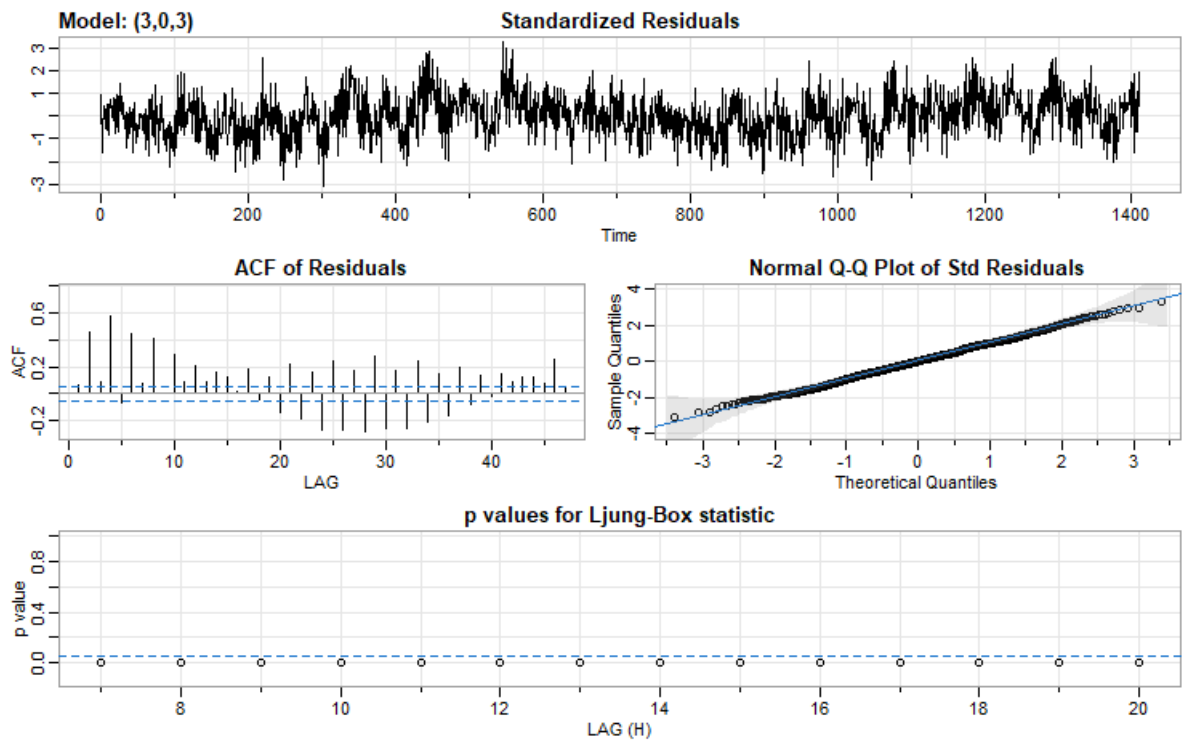


Figura 2.5: Análisis residual ARMA(3,3)

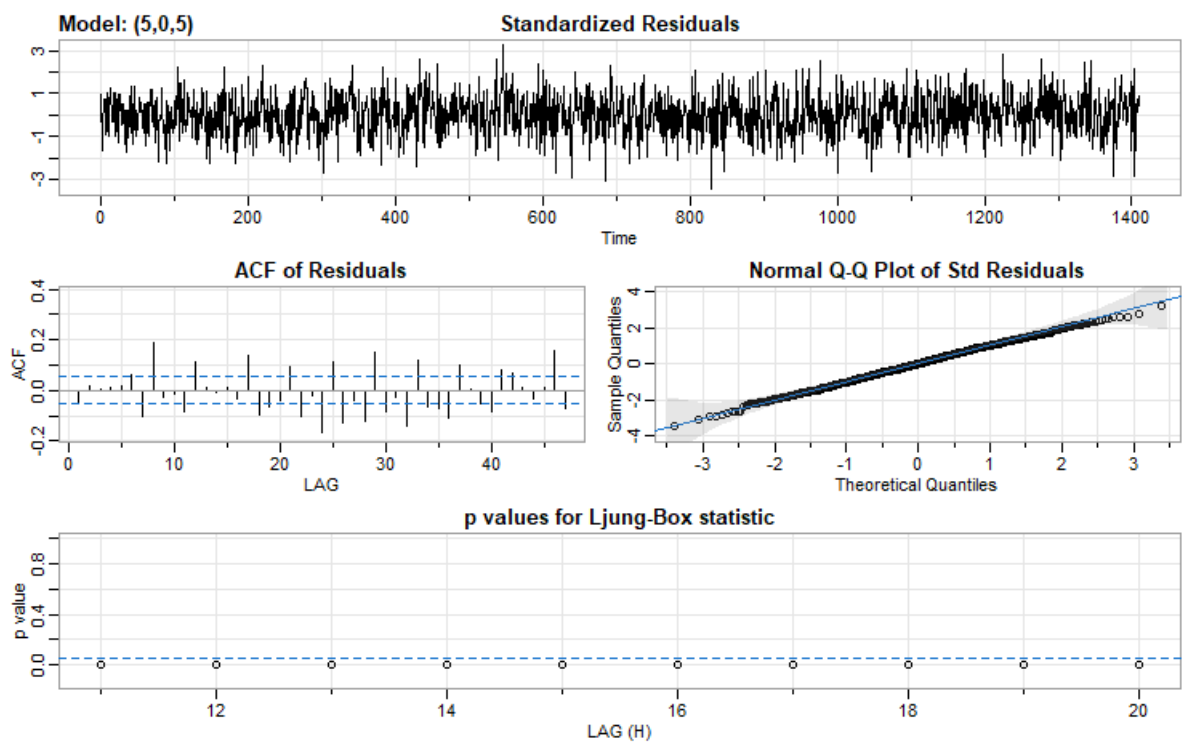


Figura 2.6: Análisis residual ARMA(5,5)

ne una variación menor y presenta una cantidad menor de rezagos sin embargo los residuos no son independientes, además los residuos estan-

darizados presentan normalidad.

Mediante el software R, utilizando la función *auto.arima* genera el mejor modelo en base a la interpretación del software; se generó el modelo ARMA(2,2), siendo uno de los modelos que comparamos (ver cuadro 2.2), ahora analizaremos los residuales del modelo ARMA(2,2) se presenta la figura 2.7

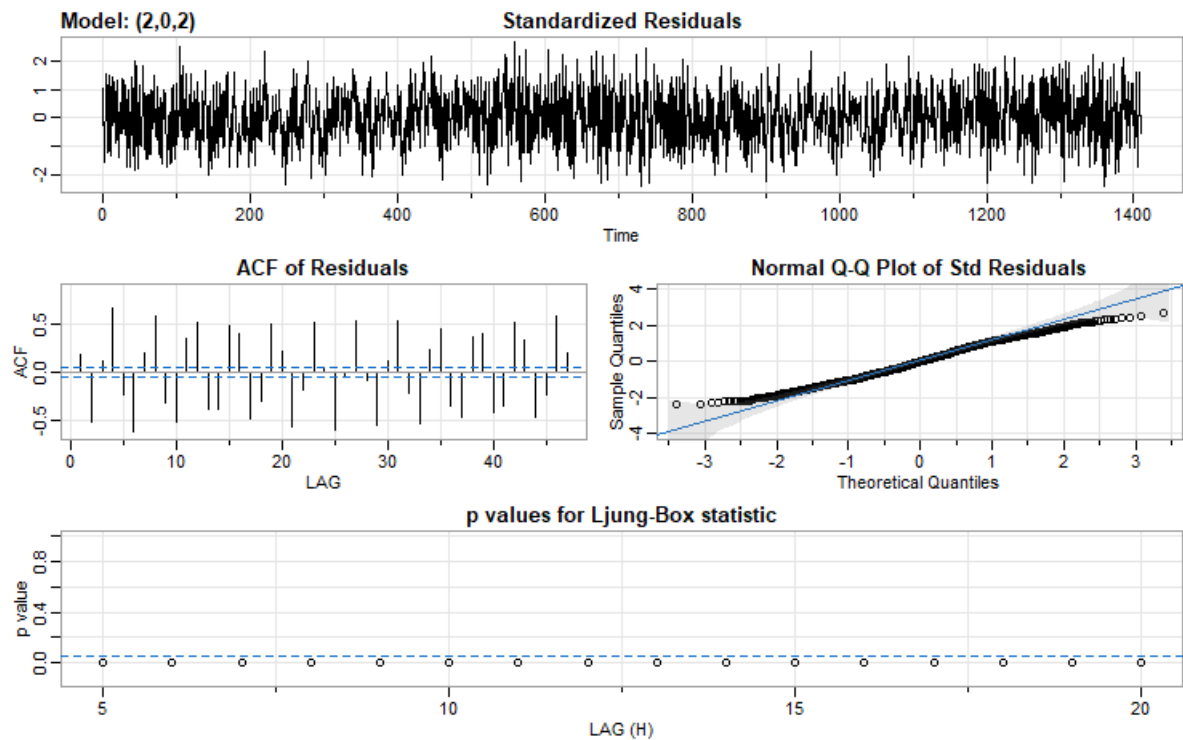


Figura 2.7: Análisis residual ARMA(2,2)

Notemos que la variación, en comparación con los residuales de los dos modelos anteriores analizados, es mayor, además los residuos no son independientes y respecto a la normalidad de los residuos los valores se encuentran dentro del rango de normalidad, excepto unos pocos valores de la cola inferior que se encuentran en el borde del rango de normalidad. Cabe mencionar que los modelos planteados sus residuos presentan heterocedasticidad, por lo cual en base a un criterio adicional (validación cruzada) escogeremos un modelo adecuado para usar de base en el modelamiento de la volatibilidad de un modelo *GARCH*.

2.5.3. Validación Cruzada

Para la validación cruzada, la base de entrenamiento comprende los 11 meses del año 2021 (enero a noviembre) y la base de prueba o test sería el último mes del año 2021 (diciembre). El siguiente cuadro 2.3 contiene los valores tanto para la base de entrenamiento como test con sus respectivas métricas.

MODELO		RMSE	MAE	MASE	U2
ARMA(2,2)	Test set	0.306	0.252	0.124	0.191
ARMA(3,3)	Test set	0.111	0.098	0.048	0.058
ARMA(3,0)	Test set	0.139	0.117	0.058	0.067
ARMA(5,0)	Test set	0.134	0.110	0.054	0.067
ARMA(5,5)	Test set	0.164	0.113	0.056	0.120

Cuadro 2.3: Validación Cruzada de los modelos ARMA

En primer lugar, señalemos que en términos de **U2** el modelo $ARMA(2,2)$ es mejor que los demás modelos planteados, sin embargo en términos de **RMSE** (raíz del error cuadrático medio) el valor del modelo $ARMA(3,3)$ está mucho más cerca al cero en comparación al modelo $ARMA(2,2)$, por lo cual presenta un sesgo menor en la predicción. También en las demás métricas, en su mayoría, el modelo $ARMA(3,3)$ tiene mejor rendimiento. Sin embargo, en base al análisis residual no es un modelo adecuado, ya que sus residuos presentan heterocedasticidad.

2.5.4. Modelo GARCH

En la figura 2.5, observamos que nuestros residuos no son independientes para el modelo $ARMA(3,3)$, siendo el que tiene mejores propiedades, calcularemos los residuales al cuadrado para verificar si hay heterocedasticidad.

Podemos observar que la varianza no es constante, ahora realizaremos la regresión de los residuos al cuadrado rezagados para concluir si tiende a ser un modelo ARCH.

Para esto contrastaremos la hipótesis nula: No presenta efectos ARCH si p -valor es mayor que 0,05, contra la hipótesis alternativa: Si presenta efectos ARCH si p -valor es menor que 0,05.

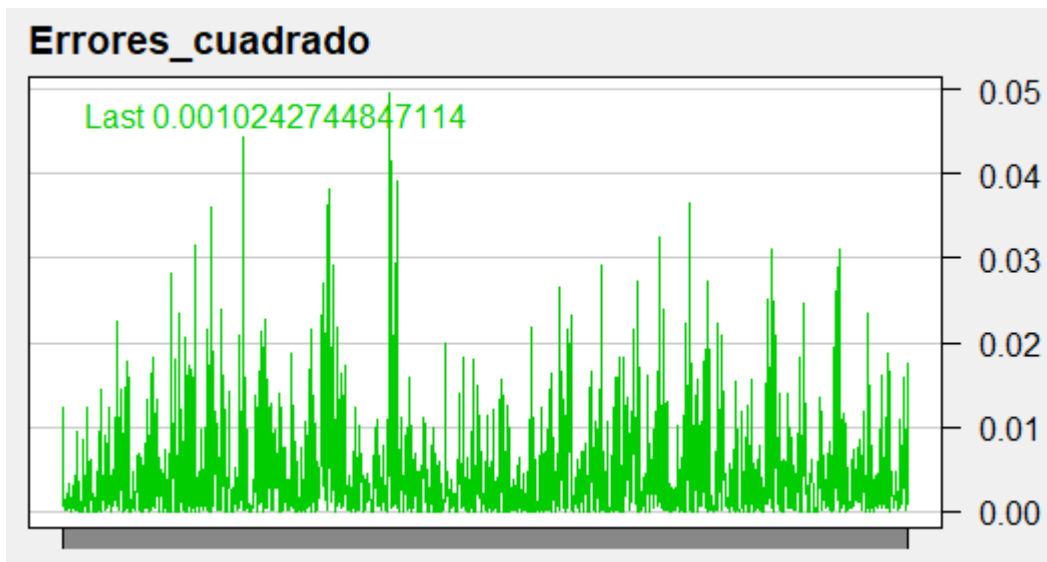


Figura 2.8: Residuales al cuadrado del modelo $ARMA(3,3)$

Por lo cual mediante el software R, se obtuvo un $p - valor = 1,7e - 05 < 0,05$. Rechazamos la hipótesis nula y concluimos que si presenta efectos ARCH en la varianza residual al cuadrado de los rezagos. En base a la Autocorrelación (ver figura 2.9) y Autocorrelación Parcial (ver figura 2.10) del modelo seleccionado, $ARMA(3,3)$ comprobaremos que la varianza es heterocedástica.

Podemos observar que tanto la autocorrelación y autocorrelación parcial salen de los límites de criterio, por lo cual presentaría ruido blanco, entonces la varianza presenta heterocedasticidad. Por lo cual generaremos un modelo GARCH, bajo la condición que el error actual se relaciona con los errores cuadrados de las innovaciones.

Partiendo del modelo con mejor rendimiento, $ARMA(3,3)$ modelamos la volatilidad de las alturas de las olas, mediante el software R y la función *Garchauto* el mejor modelo es un GARCH(1,1). A continuación mostraremos la varianza del modelo $ARMA(3,3)$

Observemos que tanto la varianza como sus residuales al cuadrado mejoran significativamente (ver figura 2.11 y figura 2.12), en casi todo el período de estudio se mantiene cercano a cero, respecto a los generados en el modelo $ARMA(3,3)$ original, mediante el modelo de la volatilidad de la altura de las olas.

Para una mejor estimación y verificación del mejor modelo es un

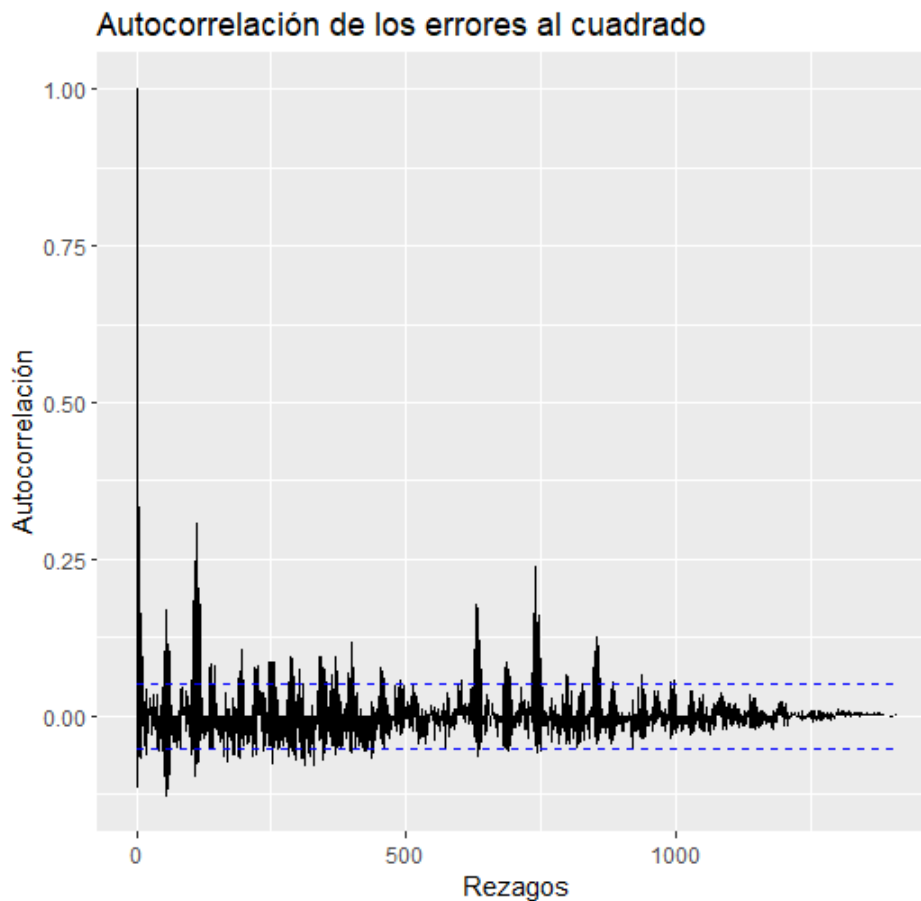


Figura 2.9: Autocorrelación de Errores al cuadrado

$GARCH(1,1)$ basado en un $ARMA(3,3)$, se realizó la estimación de los modelos $GARCH(1,1)$ basado en los dos modelos analizados con anterioridad, $ARMA(2,2)$ y $ARMA(5,5)$

Comparando la exactitud en las predicciones de cada modelo mediante la técnica de validación cruzada y tomando las métricas MSE (Error cuadrático medio) y MAE (Error absoluto medio), los resultados se presentan en el cuadro 2.4. Observemos que los resultados de MSE del modelo basado en un $ARMA(3,3)$ está más cerca al cero que los demás modelos. Cabe mencionar que el modelo basado en $ARMA(5,5)$ no presenta una diferencia tan amplia sin embargo el hecho de tener que estimar tantos parámetros genera un costo computacional más elevado y el modelo basado en un $ARMA(2,2)$ si tiene diferencias significativas en sus métricas, además en sus residuales y varianzas presenta una mayor volatilidad.

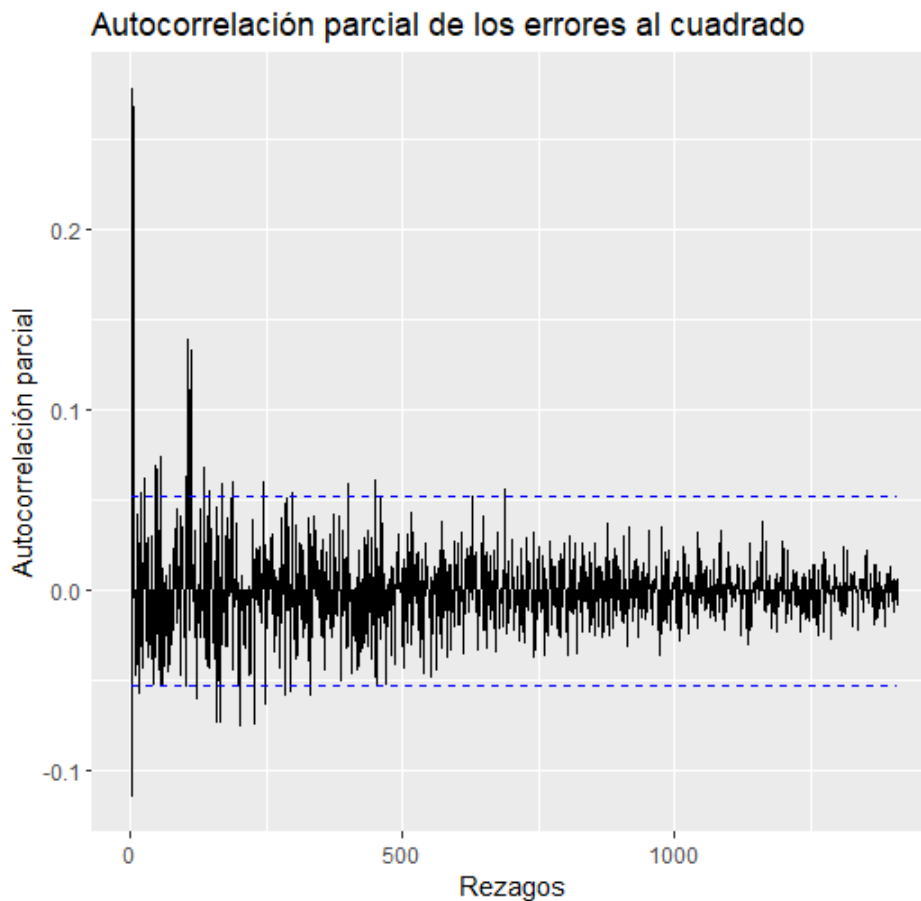


Figura 2.10: Autocorrelación parcial de Errores al cuadrado

Modelo GARCH basado	MSE	MAE
<i>ARMA</i> (3, 3)	0.012	0.092
<i>ARMA</i> (5, 5)	0.030	0.118
<i>ARMA</i> (2, 2)	0.092	0.249

Cuadro 2.4: Resumen Modelo GARCH

2.6. Cadenas de Markov

Se analizará para el mes de enero y el procedimiento se repetirá para los demás meses del año 2021.

2.6.1. Enero

Mediante Cadenas de Markov, mostraremos el mes de Enero del 2021 ver figura 2.13. Notemos que los datos de altura de las olas estan entre un valor máximo de $3m$ y mínimo de $0m$. Cumple la propiedad markoviana.

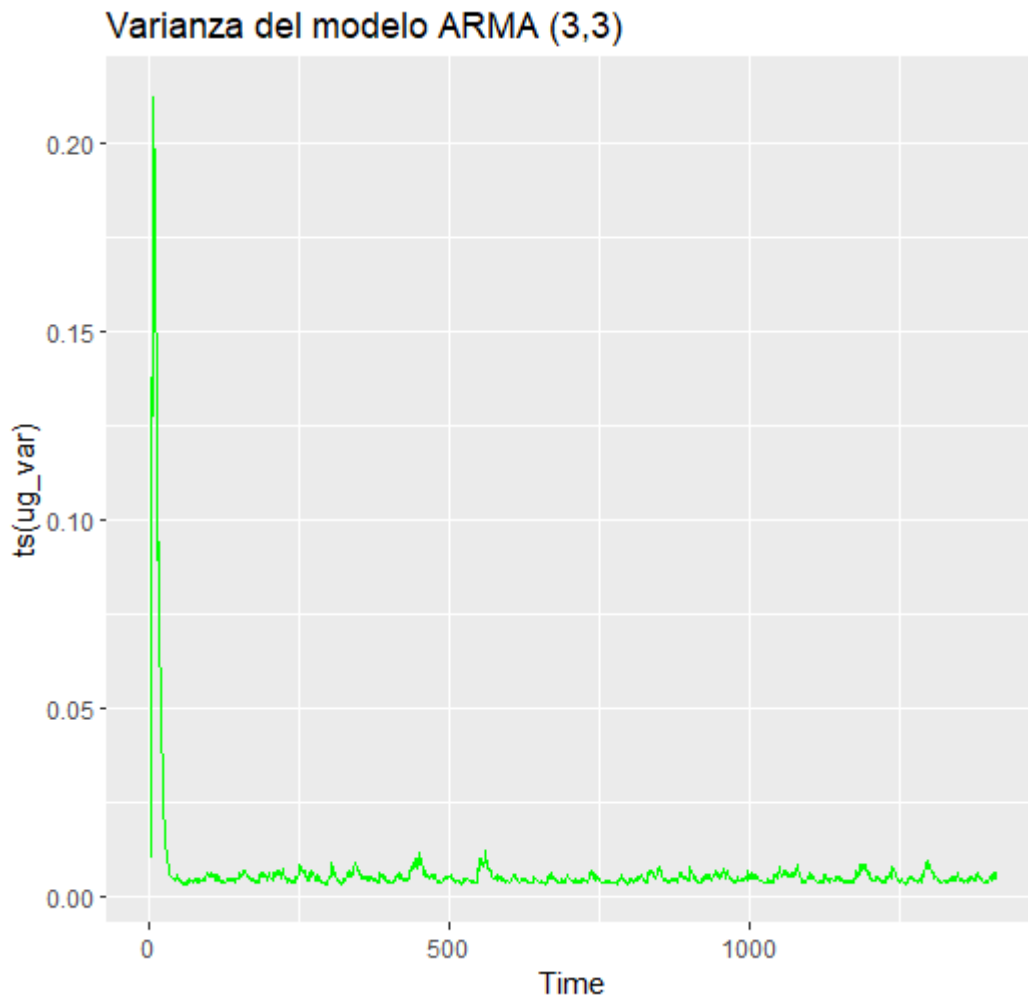


Figura 2.11: Varianza del modelo ARMA(3,3)

Los resultados se presentan a continuación:

Además en el gráfico se incluye los valores entre los cuales esta cada estado definido en la Escala Douglas. Para realizar la cadena de markov en base a la escala definida. Para esto comprobaremos si los datos del mes de Enero cumplen con la propiedad markoviana, los resultados se presentan en el anexo ???. Como el p-valor es $0,7022 > 0,05$ entonces no rechazamos la hipótesis nula, es decir la secuencia en Enero sigue la propiedad markoviana.

2.6.2. Probabilidad de Tansición $P_{i,j}$

La probabilidad de que la altura de la ola pase de un estado i (altura de la ola) a un estado j (altura de la ola), se denomina probabilidad de

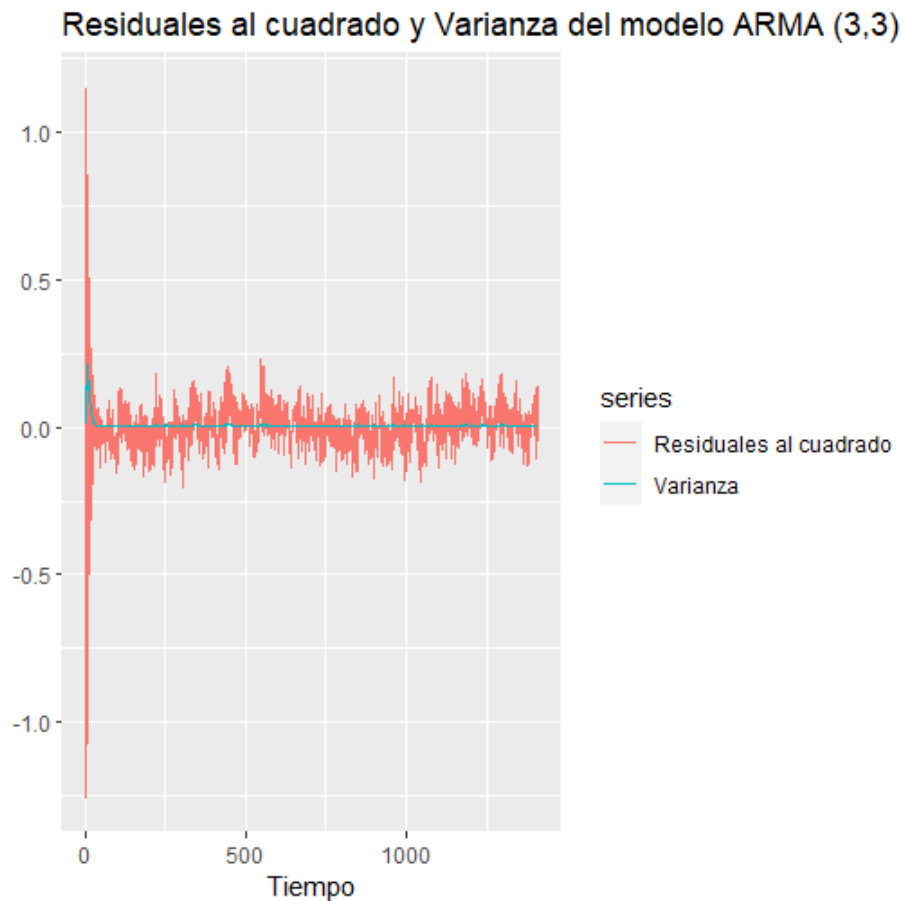


Figura 2.12: Residuales al cuadrado y Varianza del modelo ARMA(3,3)

transición de un paso y se denota $P_{i,j}$. En los casos donde la altura de la ola no pueda pasar de un estado i a un estado j la probabilidad de transición es $P_{i,j} = 0$, caso contrario si la altura de la ola sólo puede pasar de un estado i al estado j la probabilidad de transición es $P_{i,j} = 1$ [8]. Para los demás casos nos ayudaremos del estimador

$$P_{i,j} = n_{i,j}/n_i \quad (2.1)$$

donde $n_{i,j}$: son el número de veces que la altura de la ola paso desde el estado i al estado j , y n_i : son el número de veces que la altura de la ola salió del estado i .

La cadena de markov y sus probabilidades de transición entre estados, se presentan a continuación 2.14:

Notemos que entre los nodos o estados se encuentra la probabilidad

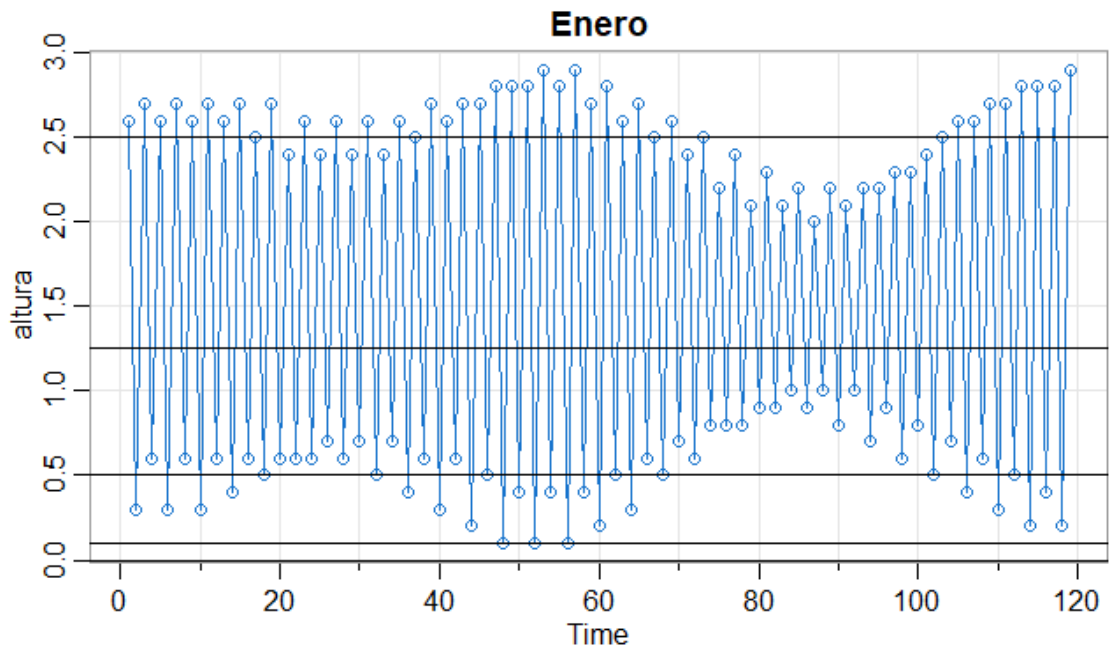


Figura 2.13: Alturas del mes de Enero

de pasar de un estado a otro o probabilidad de transición. Por ejemplo la probabilidad de pasar del estado 2, oleaje bajo, al estado 5, oleaje grueso moderado, es 0.60. Para obtener estas probabilidades se baso en la ecuacion (2.1), además se lo realiza en base a una tabla de frecuencias entre estados (ver cuadro 2.5) y posterior calcular su probabilidad de transición por filas (ver cuadro 2.6). Por ejemplo, para calcular la probabilidad de transición del estado 2 al 5 se tiene:

$$P(\text{Estado 2 al Estado 5}) = \frac{21}{35} = 0,60$$

y bajo el mismo razonamiento se calculan las demás probabilidades de transición entre los demás estados.

Estados	0	1	2	3	4	5	Total
0	0	0	0	0	0	0	0
1	0	0	0	0	0	3	3
2	0	0	0	0	3	21	24
3	0	0	0	0	21	11	32
4	0	0	3	21	0	0	24
5	0	3	21	11	0	0	35

Cuadro 2.5: tabla de frecuencias

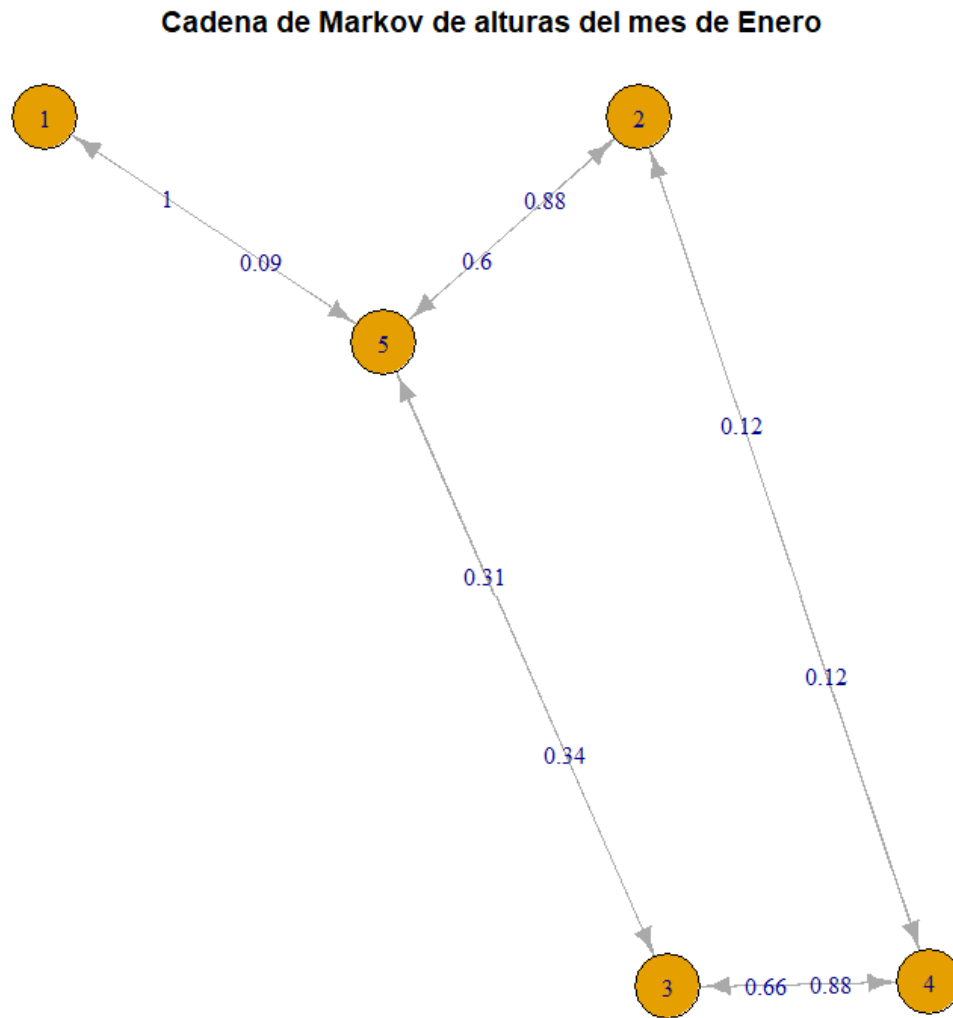


Figura 2.14: Cadena de markov de alturas del mes de enero

Mediante el software R y la función *steadyStates* podemos calcular las probabilidades de transición a largo plazo o la distribución estacionaria para el mes de Enero, los resultados se presentan en el cuadro 2.7.

A largo plazo se observa que la altura de las olas estaría en un estado 1, oleaje muy bajo, con una probabilidad de 0.025, en un estado 2, oleaje bajo, con una probabilidad de 0.203, en un estado 3, oleaje ligero, con una probabilidad de 0.271, en un estado 4, oleaje Moderado, con una probabilidad de 0.203 y en un estado 5, oleaje grueso moderado, con una probabilidad de 0.297.

Para calcular el número esperado de veces que estará la ola en cada

Estados	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	0	0	0	0	1
2	0	0	0	0	0.125	0.875
3	0	0	0	0	0.625	0.344
4	0	0	0.125	0.875	0	0
5	0	0.086	0.60	0.314	0	0

Cuadro 2.6: Matriz de transición

Estado	1	2	3	4	5
Probabilidad	0.025	0.203	0.271	0.203	0.297

Cuadro 2.7: Distribución Estacionaria

uno de los estados se puede estimar mediante:

$$E = N * p_i \quad (2.2)$$

donde N : el número de días del mes a estudiar y p_i : la probabilidad estacionaria en cada estado i . Por lo tanto el número esperado, en enero, de veces que estará la altura de la ola en cada uno de los estados se presenta en el cuadro 2.8.

Estado	1	2	3	4	5
Probabilidad	0.025	0.203	0.271	0.203	0.297
Número esperado	0.775	6.293	8.401	6.293	9.207

Cuadro 2.8: Distribución Estacionaria

2.6.3. Año 2021

Mediante el mismo razonamiento para cada mes se obtuvo los siguientes resultados resumidos en el cuadro 2.9, donde podemos observar que existen meses que cumplen con la propiedad markoviana y otros no, por lo cual para todo el año 2021 la propiedad markoviana no se cumpliría, cabe mencionar que los meses que cumplen la propiedad están constituidos de tres meses consecutivos, por ejemplo: junio, julio y agosto, así como los meses que no cumplen con la propiedad, por ejemplo: septiembre, octubre y noviembre con la particularidad del mes de abril.

Ecuador por su ubicación geográfica posee dos estaciones: invierno

Mes	p-valor	Propiedad de Markov
Enero	0.7022	Cumple
Febrero	0.1300	Cumple
Marzo	2.65e-06	No cumple
Abril	0.0651	Cumple
Mayo	0.0046	No cumple
Junio	0.1232	Cumple
Julio	0.8374	Cumple
Agosto	0.9858	Cumple
Septiembre	0.0024	No cumple
Octubre	0.0082	No cumple
Noviembre	0.0008	No cumple
Diciembre	0.1787	Cumple

Cuadro 2.9: Resumen por meses propiedad markoviana

(diciembre a mayo) y verano (junio a noviembre), para la comprobación de la propiedad de markov en cada estación, se tiene el siguiente cuadro 2.10:

Temporada	p-valor	Propiedad de Markov
Invierno: Enero - Mayo	1.908e-08	No cumple
Verano: Junio - Noviembre	1.874e-05	No cumple
Todo el año: Enero - Diciembre	0	No cumple

Cuadro 2.10: Resumen por temporada de la propiedad markoviana

Para cada estación no cumplen con la propiedad markoviana, sin embargo con lo mencionado anteriormente, podemos deducir que la propiedad markoviana se cumple para los trimestres iniciales de cada estación (ver cuadro 2.11), por falta de los datos de altura del mes de diciembre de 2020 se realiza la comprobación solo con el mes de enero y febrero para el inicio de la estación de invierno.

Mes	p-valor	Propiedad de Markov
Enero a Febrero	0.468	Cumple
Marzo a Mayo	5.43e-08	No cumple
Junio a Agosto	0.316	Cumple
Septiembre a Noviembre	3.24e-07	No cumple

Cuadro 2.11: Resumen por trimestre de la propiedad markoviana

Entonces mediante el mismo razonamiento utilizado en el mes de Enero, calcularemos tanto las probabilidades de transición y las probabilidades estacionarias de cada trimestre que cumple con la propiedad markoviana (anexo [3.3](#)).

Capítulo 3

Resultados, conclusiones y recomendaciones

3.1. Resultados y Conclusiones

En esta sección presentaremos los resultados principales de la modelización del oleaje en el cantón Bahía de Caráquez, en las diferentes temáticas planteadas.

3.1.1. Serie de tiempo

Para determinar el mejor modelo para predecir el oleaje en el cantón Bahía de Caráquez, se realizó la comparación de 5 modelos ARMA en base los estadísticos AIC, AICc y BIC (ver cuadro 2.2), donde se obtuvo que el mejor modelo en comparación es el $ARMA(5, 5)$.

Posterior, se realizó un análisis residual de los modelos planteados (ver figuras 2.7, 2.5, 2.6). Dando como resultado que los tres modelos en la varianza presentaban heterocedasticidad. Luego, se realizó la validación cruzada de cada modelo para seleccionar un modelo, en base a las métricas analizadas del cuadro 2.3, de donde el modelo con las mejores métricas es el modelo $ARMA(3, 3)$, que nos sirve de base para modelar un modelo $GARCH$.

Realizamos la comparación de tres modelos $GARCH$ basados en los mejores modelos en predicción de los datos de altura de la ola, obtenien-

do el mejor modelo es un $GARCH(1,1)$ basado en un $ARMA(3,3)$ como se presenta en el cuadro 2.4. Consiguiente verificaremos los residuos cuadrados y varianza mejoraron como se muestra en la siguiente figura:

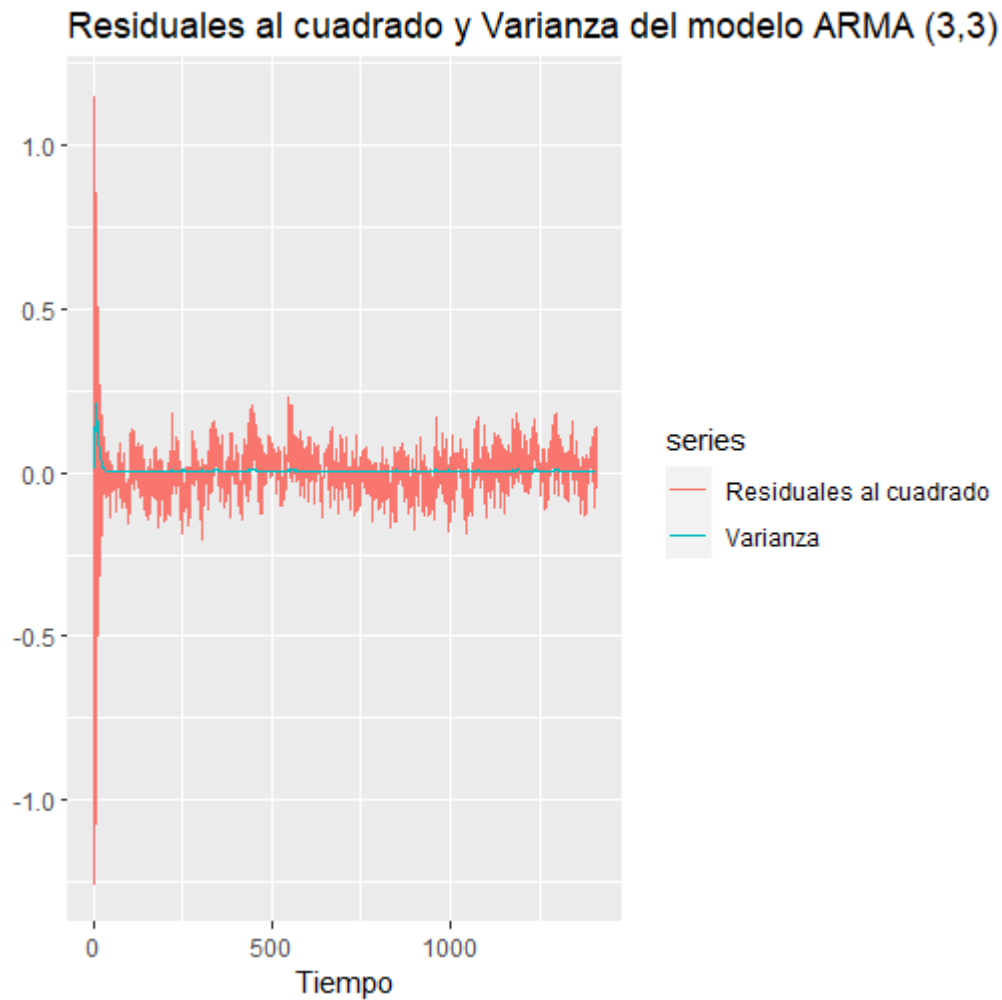


Figura 3.1: Residuales y varianza del modelo $ARMA(3,3)$

Los coeficientes del modelo son:

mu	ar1	ar2	ar3	ma1	ma2	ma3	omega	alpha1	beta1
1.55	-1.16	-1.10	-0.98	0.42	0.85	0.07	0.0006	0.098	0.768

Cuadro 3.1: Coeficientes del modelo $GARCH(1,1)$

Ahora presentaremos la predicción del modelo junto al ajuste realizado, donde observamos que la volatilidad se ve corregida en gran parte por el modelo $GARCH(1,1)$ (ver figura 3.2). Por lo tanto el modelo adecuado para realizar las predicciones de futuros oleajes en la puerto de Bahía de Caráquez es un modelo $GARCH(1,1)$ basado en un modelo $ARMA(3,3)$.

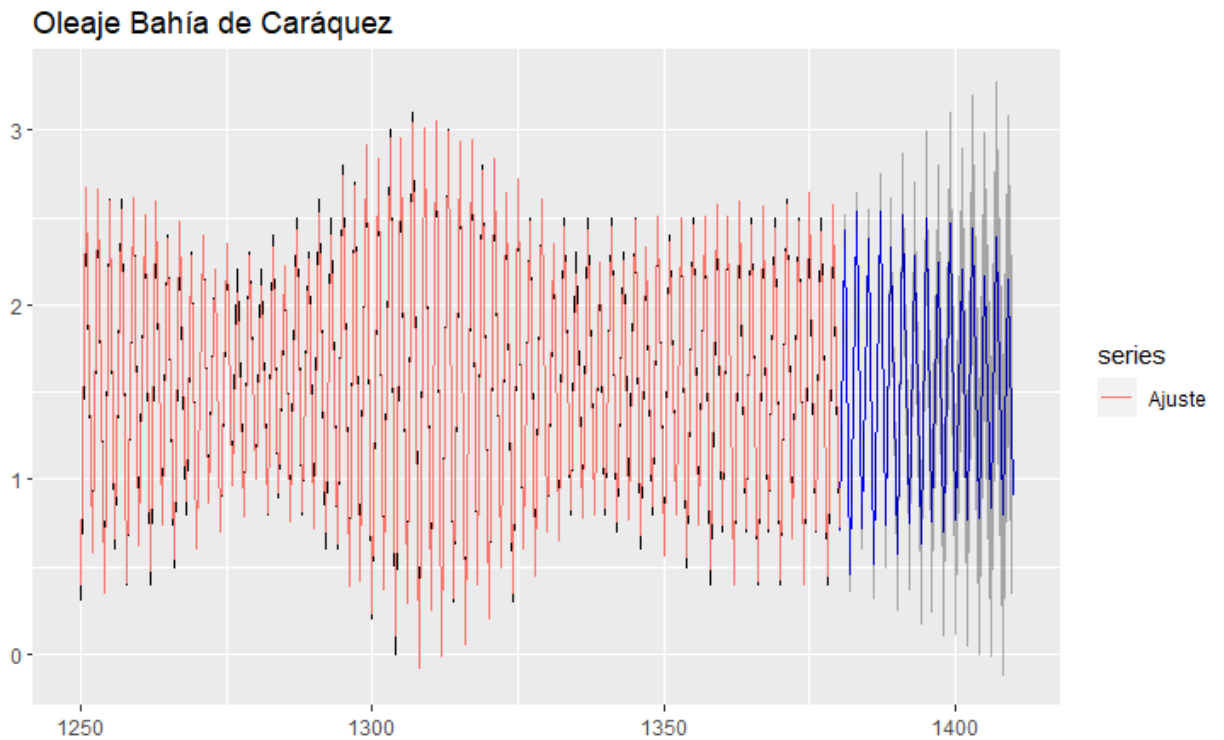


Figura 3.2: Ajuste y predicción del oleaje mediante el modelo $GARCH(1, 1)$ basado en un $ARMA(3, 3)$

3.1.2. Cadenas de Markov

Enero es un proceso Markoviano, puesto que cumple con la propiedad markoviana, se generaron buenos resultados para la estimación del valor esperado para la altura de las olas estaría en unos de los seis estados definidos, con su respectiva probabilidad estacionaria, como se muestra en el cuadro 2.8.

Se muestra que en el año 2021 no cumplía con la propiedad markoviana, por lo cual se realizó un análisis por mes (ver cuadro 2.9). Se realizó una análisis por temporada que no cumplían con la propiedad markoviana. Sin embargo, Cabe mencionar que el primer trimestre es de diciembre a febrero, pero al no contar con los datos del año 2021, entonces se procedió a realizar esta parte del estudio con los meses de enero y febrero, como se muestra en el siguiente cuadro 3.2:

Por ultimo, se estimó la probabilidad estacionaria para determinar el número esperado de veces que estará en cada estado, definido mediante

Mes	p-valor	Propiedad de Markov
Enero a Febrero	0.468	Cumple
Marzo a Mayo	5.43e-08	No cumple
Junio a Agosto	0.316	Cumple
Septiembre a Noviembre	3.24e-07	No cumple

Cuadro 3.2: Resumen por trimestre de la propiedad markoviana

la escala douglas, como se muestra en el siguiente cuadro:

Estado	1	2	3	4	5
Probabilidad	0.017	0.2404	0.2404	0.2205	0.279
Número esperado	1.003	14.18	14.18	13.01	16.46

Cuadro 3.3: Probabilidades estacionarias y Número esperado - Primer trimestre

y se replicó con el tercer trimestre que cumplía con la propiedad markoviana ??.

3.2. Recomendaciones

1. Para ampliar el estudio, se requiere de datos más accesibles que permitan una mejor manipulación de los mismos, junto a más factores asociados a cada altura de la ola.
2. En un trabajo futuro el análisis temporal, se lo amplía mediante redes neuronales, siendo un método más sofisticado que nos daría resultados más aproximados y a su vez disminuiría el error de predicción en el modelo que se generaría.
3. En procesos estocásticos, referente a cadenas de markov se limita a realizar de forma trimestral para una mejor aproximación del método se necesitaría involucrar variables o factores adicionales para una mejor estimación de parámetros y a su vez predicciones.

Referencias bibliográficas

- [1] D.A.Pierce Box. G.E.P. Distributions of residual autocorrelations in autoregressive integrated moving average models. 1970.
- [2] B. Chen. A new method to estimate wave height of specified return period. *Oceanol Limmol*, 2017.
- [3] Richard Durrett. *Essential of Stochastic Processes*. Springer Texts in Statistics, 2012.
- [4] F. Gonzalez. Modelización estadística de eventos extremos de oleaje y nivel de mar. 2013.
- [5] INOCAR. tabla-mareas. <https://www.inocar.mil.ec/web/index.php/productos/tabla-mareas>, 2018.
- [6] I love PDF. Convierte pdf a excel. https://www.ilovepdf.com/es/pdf_a_excel, 2018.
- [7] Wang L Lui G, Chen B. Wave height statistical characteristic analysis. *Oceanol Limmol*, 2019.
- [8] Ferndado Luis. Predicciones de markov aplicadas en el programa de ingeniería industrial de la universidad nacional experimental de táchira. *Actualidad y nuevas tendencias*, I(2), enero - junio 2019.
- [9] Networking Meteorology. Douglas scale. <https://www.meteorologiaenred.com/en/douglas-scale.html>., 2018.
- [10] David S. stoner Robert H. Shumway. *Time series Analysis and its Applications*. Springer Texts in Statistics, 2010.

- [11] D. Zhao F. Wu. Calculations on stopping time and return period.
Springer Nature, 2020.