



ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

MODELIZACIÓN DE MAREJADAS EN ESMERALDAS

PROYECTO DE INVESTIGACIÓN

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

LUIS FABIAN AMAGUA OBANDO

luis.amagua@epn.edu.ec

DIRECTOR: MSC. YANDIRA DENISSE CUVERO CALERO

yandira.cuvero@epn.edu.ec

DMQ, FEBRERO 2022

CERTIFICACIONES

Yo, LUIS FABIAN AMAGUA OBANDO, 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.

Luis Fabian Amagua Obando

Certifico que el presente trabajo de integración curricular fue desarrollado por Luis Fabian Amagua Obando, bajo mi supervisión.

Msc. 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 el(los) producto(s) resultante(s) del mismo, es(son) público(s) y estará(n) 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.

Luis Fabian Amagua Obando

Msc. Yandira Denisse Cuvero Calero

RESUMEN

Este trabajo identifica la distribución que permite describir los datos de altura máxima diaria del oleaje en el cantón Esmeraldas. La teoría de valores extremos permite encontrar una distribución que describa el oleaje, siendo las distribuciones Normal y Gamma de dos parámetros las que presentan mejores ajustes en base a los datos descargados de la página del INOCAR.

Para determinar la distribución, se analiza las distribuciones de valores extremos, así como las distribuciones Normal y Gamma. Se concluye que las distribuciones truncadas permiten modelar de manera más eficiente el oleaje, si este se separa en las estaciones de verano e invierno, si se considera que la altura máxima del oleaje se encuentra en grado cinco según la escala Douglas, como el límite de truncamiento superior.

La esperanza del tiempo de parada representa el período de retorno en hidrología, se demuestra que el tiempo de parada $N = \inf\{n \geq 1 : X_n > x_0\}$ sigue una distribución geométrica y se analiza la relación entre el período de retorno y la altura del oleaje.

Palabras clave: Distribuciones truncadas, periodo de retorno oleaje.

ABSTRACT

This work identifies the distribution that allows describing the maximum daily surge height data in the Esmeraldas canton. The extreme value theory allows finding a distribution that describes the surge, being the two-parameter Normal and Gamma distributions the best fits based on the data downloaded from the INOCAR web page.

To determine the distribution, the extreme value distributions are analyzed, as well as the Normal and Gamma distributions. It is concluded that the truncated distributions allow to model more efficiently the surge, if it is separated into summer and winter seasons, if the maximum height of the surge is considered to be at grade five according to the Douglas scale, as the upper truncation limit.

The stopping time expectation represents the return period in hydrology, it is shown that the stopping time $N = \inf\{n \geq 1 : X_n > x_0\}$ follows a geometric distribution and the relationship between return period and surge height is analyzed.

Keywords: Truncated distribution, return period surge.

AGRADECIMIENTOS

A mi madre Melba y mi hermano Jhon por el apoyo incondicional.

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

A todos los profesores de la Facultad de Ciencias por la enseñanza académica, profesional y moral.

A mis amigos, principalmente a Katherine M. y Jhosselyne O. por los consejos y recomendaciones para no rendirme.

DEDICATORIA

*A mi madre Melba por el cariño
y apoyo para seguir mis sueños.*

Índice general

1. Descripción del componente desarrollado	1
1.1. Objetivo general	3
1.2. Objetivos específicos	3
1.3. Alcance	3
1.4. Marco teórico	3
1.4.1. Distribuciones de Valores extremos	4
1.4.2. Distribución Gamma	5
1.4.3. Distribuciones truncadas	6
1.4.4. Método de máxima verosimilitud	7
1.4.5. Tiempo de parada	8
1.4.6. Período de retorno	9
2. Metodología	10
2.1. Región de estudio	10
2.2. El INOCAR	11
2.2.1. La marea	11
2.2.2. Escala Douglas	12
2.3. Tratamiento de datos	13
2.4. Modelización	17
2.4.1. Distribuciones de probabilidad	17

2.5. Comparación de los modelos	19
2.5.1. Estimación de los parámetros	20
2.6. Distribuciones truncadas	22
2.6.1. Distribución Normal Truncada	23
2.6.2. Distribución Gamma truncada	26
2.7. Período de retorno	29
2.8. Aplicativo Shiny	31
3. Resultados, conclusiones y recomendaciones	33
3.1. Resultados y Conclusiones	33
3.1.1. Para Invierno	33
3.1.2. Para Verano	36
3.2. Recomendaciones	38
A. Anexos	39
A.1. Teorema tiempo de parada	39
A.2. Estadística Descriptiva	41
A.3. Distribuciones para verano	42
A.4. Pruebas de hipótesis verano	43
A.5. Distribución Normal verano	44
A.6. Distribución Gamma verano	44
A.7. Distribución Normal truncadas verano	45
A.8. Distribución Gamma truncada	46
A.9. EMV de la distribución Normal	47
A.10. EMV de la distribución Gamma truncada	47
Bibliografía	48

Índice de figuras

2.1. Pleamar y bajamar	11
2.2. Marea enero	14
2.3. $H_{\text{máx}}$ oleaje en Esmeraldas	15
2.4. Diagrama de cajas invierno	16
2.5. Gráfico de Cullen y Frey en invierno	17
2.6. Distribuciones invierno	18
2.7. Distribuciones acumulada en invierno	18
2.8. Distribución Normal para $H_{\text{máx}}$ en invierno.	20
2.9. Distribución Gamma para $H_{\text{máx}}$ en invierno.	21
2.10. Distribuciones Normales truncadas en invierno	24
2.11. Distribuciones Gamma truncadas en invierno	27
2.12. Aplicativo Shiny: Distribuciones truncadas	32
2.13. Aplicativo Shiny: Estadística Básica	32
3.1. Altura vs periodo de retorno	35
3.2. Altura vs periodo de retorno	36
3.3. Altura vs periodo de retorno en verano.	37
3.4. Altura vs periodo de retorno en verano.	38
A.1. Diagrama de cajas verano	41

A.2. Gráfico de Cullen y Frey en verano	42
A.3. Distribuciones verano	42
A.4. Distribuciones acumulada en verano	43
A.5. Distribuciones para $H_{\text{máx}}$ en verano.	44
A.6. Distribución Gamma para $H_{\text{máx}}$ en verano.	44
A.7. Distribuciones Normales truncadas en verano	45
A.8. Distribuciones Gamma truncadas en verano	46

Capítulo 1

Descripción del componente desarrollado

Establecer proyectos de protección como rompeolas y barreras contra inundaciones es la clave para prevenir las catástrofes marinas causadas por las condiciones extremas del mar, determinar la distribución del oleaje es de gran utilidad para diseño y correcta construcción de puertos, malecones, o el fortalecimiento de la infraestructura y de esta forma evitar daños materiales [7].

La marejada se define como el aumento anormal de agua [5], considerando la escala Douglas [12] se puede clasificar el oleaje de acuerdo al nivel de la altura de las olas desde el grado cero hasta el grado nueve, en el cual la altura de las olas que superan los 14[m]. Se considera que el oleaje en Esmeraldas tiene como máximo grado cinco debido a que la altura máxima de las olas se encuentra por debajo de 3.6[m].

En época de invierno se presentan oleajes fuertes provenientes del Pacífico norte que representa pérdidas tanto naturales como económicas. La marejada combinada con otros fenómenos como el del Niño o fuertes aguajes, dejan pérdidas significativas para sus poblaciones [7].

Esmeraldas está ubicada en la costa norte de Ecuador, en una zona con alta probabilidad de ocurrencia de terremoto tsunami-génicos [10], es vulnerable a: marejadas, fuertes marejadas y a un oleaje grueso moderado; lo cual puede causar grandes daños materiales. Los daños causados por marejadas en la provincia de Esmeraldas dejan grandes pérdidas eco-

nómicas y materiales como sucedió en 2018, en el cual varios cantones estuvieron en emergencia [17].

El modelo de ola clásico representado por Longuet-Higgins en 1957 muestra que este es un proceso estocástico debido a que es un fenómeno aleatorio, cuya dinámica se rige por el azar en el tiempo [2]. La teoría de valores extremos proporciona una base para modelar el oleaje [11], permite conocer el máximo oleaje en una determinada región, para estimar el nivel de las mareas y la altura de las olas de diferentes períodos de retorno. La distribución Gumbel y la distribución Weibull se usan con frecuencia para hacer el modelado de condiciones extremas de mar [14][13], además se consideran las distribución Normal doblemente truncada y la distribución Gamma truncada por derecha [4] [3].

Se desarrolló un aplicativo que permite encontrar la función de distribución que mejor se ajuste a los datos del oleaje del puerto de Esmeraldas, el cual puede ser utilizado como referencia para la construcción de muros de contención, rompe olas, puertos pesqueros y prevenir pérdidas materiales o humanas. En Ecuador los datos de las alturas de las olas son obtenidos por los mareógrafos del INOCAR [18], el aplicativo en Shiny carga los datos de las tablas de mareas del INOCAR y obtiene los parámetros de las distribuciones mencionadas anteriormente y las pruebas de hipótesis para verificar si se ajustan a las distribuciones.

1.1. Objetivo general

Modelización de marejadas en Esmeraldas.

1.2. Objetivos específicos

1. Comprender el uso de distribuciones estadísticas para el modelamiento del oleaje.
2. Estudiar el tiempo de parada en marejadas.
3. Realizar el modelamiento de la altura del oleaje mediante el uso de un software libre.

1.3. Alcance

Se analizan las distribuciones que se ajustan a los datos de la máxima altura diaria del oleaje tomados en las estaciones de verano e invierno desde enero hasta noviembre de 2021 en el cantón Esmeraldas. Además, se comparan los resultados de las distribuciones que se ajustan con sus distribuciones truncadas y seleccionar el mejor ajuste entre ellas.

1.4. Marco teórico

Las distribuciones de probabilidad comúnmente utilizadas en hidrología son la distribución Normal; las distribuciones de Gumbel y Weibull que son parte de las distribuciones de valores extremos [16], la distribución Pearson III; además la distribución Gamma que tiene la ventaja de tomar valores no negativos [1].

Las distribuciones usuales pueden tomar valores fuera del rango de interés, impidiendo un ajuste apropiado a los datos. Al considerar distribuciones truncadas se tiene la ventaja de que están delimitadas por izquierda o por derecha para algún valor denominado punto de truncamiento y de esta forma seleccionar un subconjunto de valores adecuados.

El determinar el tiempo de parada que corresponde al período de retorno en hidrología, permite identificar el momento en el que la altura máxima de las olas sobrepasa el nivel de retorno por primera vez y de esta forma tomar las previsiones pertinentes para evitar pérdidas futuras.

1.4.1. Distribuciones de Valores extremos

Existen varias aplicaciones de la teoría de valores extremos en la climatología, oceanografía, para la modelización de campos de viento, aplicaciones en el contexto de la ingeniería de la corrosión, entre otras [4].

El modelo que representa la teoría de valores extremos se centra en el comportamiento estadístico [8]:

$$M_n = \text{máx}\{X_1, X_2, \dots, X_n\}$$

donde $\{X_1, X_2, \dots, X_n\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas que tienen una función de distribución F y M_n representa el máximo en n unidades de tiempo. Si n es el número de observaciones en un año, entonces M_n corresponde al máximo valor anual [16].

La distribución M_n se puede derivar de todos los valores de n

$$\begin{aligned} P(M_n \leq x) &= P(X_1 \leq x, \dots, X_n \leq x) \\ &= P(X_1 \leq x) \cdots P(X_n \leq x) \\ &= [F(x)]^n \end{aligned}$$

Teorema 1 (Teorema de Fisher-Tippett-Gnedenko). Sean $\{X_1, \dots, X_n\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, sea $M_n = \text{máx}\{X_1, \dots, X_n\}$. Si existen sucesiones $\{a_n\}_{n \in \mathbb{N}}$ con $a_n > 0$ y $\{b_n\}_{n \in \mathbb{N}}$ con $b_n \in \mathbb{R}$, tal que

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x)$$

existe, entonces G es una función de distribución no degenerada.

Si el límite anterior converge, se tiene que $G(x)$ tiene la forma

$$G_\gamma(x) = \exp(-(1 + \gamma x)^{-1/\gamma}),$$

con $1 + \gamma x_{a,b} > 0$, para algún parámetro γ .

Se trata de la función de distribución acumulativa de la distribución de valores extremos generalizada (GEV) con índice de valor extremo γ y pertenece a una de las siguientes familias de distribuciones, conocidas como distribuciones de valores extremos las cuales se muestran a continuación.

▪ **Tipo I o distribución de Gumbel**

$$G(z) = \exp \left\{ -\exp \left[- \left(\frac{z-b}{a} \right) \right] \right\}, \quad z \in \mathbb{R}$$

▪ **Tipo II o distribución de Fréchet**

$$G(z) = \begin{cases} 0, & \text{si } z < b \\ \exp\{-(\frac{z-b}{a})^{-\beta}\}, & \text{si } z \geq b \end{cases}$$

▪ **Tipo III o distribución de Weibull**

$$G(z) = \begin{cases} \exp\{-(\frac{z-b}{a})^\beta\}, & \text{si } z \leq b \\ 1, & \text{si } z > b \end{cases}$$

en donde los parámetros b, a, β son los parámetros de localización, escala y forma respectivamente.

1.4.2. Distribución Gamma

Sea X una variable aleatoria continua que sigue una distribución Gamma de dos parámetros $X \sim \text{Gamma}(\alpha, \beta)$.

$$f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0, \quad \alpha, \beta > 0 \quad (1.1)$$

donde $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

Al reemplazar x por $(x - \gamma)/\beta$, donde γ es el parámetro de localización, se tiene la distribución Gamma de 3 parámetros, también conocida como distribución Pearson III.

$$f(x; \alpha, \beta, \gamma) = \frac{1}{\beta^\alpha \Gamma(\alpha)} (x - \gamma)^{\alpha-1} e^{-(x-\gamma)/\beta}, \quad x \geq \gamma.$$

1.4.3. Distribuciones truncadas

Sea X una variable aleatoria continua asociada a una función de distribución, si se excluye algún conjunto de valores en el rango de X ; es decir, tomamos un subconjunto del rango de X , entonces se dice que la distribución de probabilidad para la variable aleatoria X está truncada.

Definición 1 (Distribución truncada). *Sea X una variable aleatoria continua definida en el espacio de probabilidad (Ω, \mathcal{F}, P) y sea $T \in \mathcal{B}$ tal que $0 < P(X \in T) < 1$, donde \mathcal{B} es una σ -álgebra del conjunto de números reales R , entonces la distribución condicional $P(X \leq x \mid X \in T)$ definida para cualquier real x , se llama la distribución truncada de X .*

Sean $f_X(x)$ y $F_X(x)$ la función de densidad de probabilidad y la función de distribución acumulativa respectivamente de la variable aleatoria continua X . Si la variable aleatoria con la función de distribución truncada $P(X \leq x \mid X \in T)$ se denota por Y , entonces la función de distribución acumulativa $G(y)$, y la función de densidad de probabilidad $g(y)$ para la variable aleatoria Y están dadas por

$$G_Y(y) = P(Y \leq y \mid Y \in T) = \frac{P(Y \leq y, Y \in T)}{P(Y \in T)} = \frac{\int_{(-\infty, y] \cap T} f_X(u) du}{\int_T f_X(u) du}$$

y la función de densidad esta dada por

$$g_Y(y) = \begin{cases} \frac{f_X(y)}{\int_T f_X(u) du}, & \text{si } y \in T \\ 0, & \text{si } y \notin T \end{cases} \quad (1.2)$$

hPara analizar la distribución truncada se debe tomar el conjunto T

de forma adecuada, se tienen 3 posibles casos.[3]

- **Truncada por la izquierda**

En la distribución truncada inferiormente se han omitido las observaciones por debajo de algún punto de truncamiento a , entonces el conjunto de truncamiento es $T = [a, \infty)$. Para el estudio del oleaje se conoce que la altura de las olas es no negativa, entonces como se menciona en [2] se puede considerar el conjunto $T = (0, \infty)$.

- **Truncada por la derecha**

En la distribución truncada superiormente se han omitido las observaciones por encima de algún punto de truncamiento b , es decir el conjunto de truncamiento es $T = (-\infty, b]$. Para el estudio tener alturas de oleaje negativas no es una realidad, entonces no es considerado este caso en la distribución truncada.

- **Doblemente truncada**

En la distribución doblemente truncada se considera solamente la parte media de todas las observaciones; es decir, en conjunto de truncamiento es $T = [a, b]$. En el conjunto de datos disponibles se puede considerar esta distribución en donde el truncamiento por la izquierda es cero, y para el truncamiento por la derecha de debe determinar el valor del umbral.

1.4.4. Método de máxima verosimilitud

El método de máxima verosimilitud consiste en maximizar la función de verosimilitud para que la distribución de los datos se aproxime a la distribución teórica. Los estimadores de máxima verosimilitud son los principales estimadores para calcular las estimaciones de los parámetros de la distribución a partir de muestras truncadas [3].

En este trabajo se utilizó el estimador de máxima verosimilitud para obtener los parámetros estimados de las distribuciones: Normal, Gamma, Gumbel y Weibull.

Definición 2. Sea X_1, \dots, X_n una sucesión de variables aleatorias independientes e idénticamente distribuidas con función de densidad $f(x; \theta)$. La

función de verosimilitud $\mathcal{L}_n(\theta)$ se define como:

$$\mathcal{L}_n(\theta) = f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

El método de máxima verosimilitud (EMV) consiste en obtener el valor de θ de tal forma que maximice la función de verosimilitud $\mathcal{L}_n(\theta)$.

Las propiedades principales para los estimadores de máxima verosimilitud son las siguientes:

- Consistencia: $\hat{\theta}_n \xrightarrow{p} \theta$, donde θ es el verdadero valor del parámetro.
- Equivariante: Si $\hat{\theta}$ es el EMV de θ , entonces $g(\hat{\theta})$ es el EMV de $g(\theta)$.
- Asintóticamente normal: $\sqrt{n} \frac{\hat{\theta}_n - \theta}{\hat{se}} \xrightarrow{D} N(0, 1)$, donde \hat{se} es el error estándar de $\hat{\theta}_n$.
- Asintóticamente eficiente: en el conjunto de estimadores con comportamiento estable, el EMV tiene la menor varianza para muestras grandes.

1.4.5. Tiempo de parada

Definición 3 (Filtración). Una filtración es una colección no numerable de sub σ -álgebras $\{\mathcal{A}_t\}_{t \geq 0}$ tal que $\mathcal{A}_s \subseteq \mathcal{A}_t$, cuando $0 \leq s \leq t$.

Un espacio probabilístico (Ω, \mathcal{A}, P) dotado de una filtración $\{\mathcal{A}_t\}_{t \geq 0}$ se llama espacio probabilístico filtrado y se denota por $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t \geq 0}, P)$.

Una filtración permite definir el tiempo de parada.

Definición 4 (Tiempo de Parada). Sean (Ω, \mathcal{A}) un espacio medible y $\{\mathcal{A}_t\}_{t \geq 0}$ una filtración de \mathcal{A} . Una variable aleatoria T que toma valores en $\{0, 1, 2, \dots\}$ definida en (Ω, \mathcal{A}) , es un tiempo de parada con respecto a la filtración $\{\mathcal{A}_t\}_{t \geq 0}$, si $\{\omega : T(\omega) \leq t\} \in \mathcal{A}_t$ para cada $t \geq 0$.

En el sistema de previsión y alerta de catástrofes causadas por mareas cuando el nivel del agua supera un determinado valor es necesario emitir una alerta. Determinar la distribución del tiempo de parada es de importancia para la previsión y alerta de mareas [2].

Teorema 2 (Distribución del tiempo de parada.). Sean $\{X_1, X_2, \dots\}$ variables aleatorias independientes e idénticamente distribuidas que representan la altura del oleaje extrema anual en un espacio de probabilidad (Ω, \mathcal{F}, P) . Si T es el período de retorno y $N = \inf\{n \geq 1 : X_n > x_0\}$ entonces el tiempo de parada obedece una distribución geométrica [2].

La demostración se muestra en el anexo A.1.

1.4.6. Período de retorno

Definición 5 (período de retorno). Sea A un suceso y T el tiempo aleatorio que transcurre entre ocurrencias sucesivas de ese evento. Al valor medio τ , de la variable T se llama período de retorno de A , es decir que es el tiempo medio que tarda en retornar ese suceso [15]. Entonces se tiene

$$T(x) = \frac{1}{P(X > x)}$$

donde $T(x)$ es el período de retorno correspondiente a un evento de magnitud x , y $P(X > x)$ es la probabilidad de que ocurra un evento de magnitud igual o superior a x .

El período de retorno representa el número de años en que es probable que se vuelvan a producir alturas de ola extremas superiores a un valor denominado umbral, entonces si se produce en promedio dentro de 100 años este valor será el valor máximo una vez en 100 años, en consecuencia su período de retorno será de 100 años. La esperanza del tiempo de parada es el período de retorno.

El momento en que excede el nivel de retorno por primera vez es el valor del tiempo de parada. El tiempo de parada representa una anomalía en los patrones normales de la altura del oleaje, que puede indicar donde se producirá un incremento en el oleaje debido algún fenómeno natural como los tsunamis, tifones, etc, identificar estos eventos permite mejorar los preparativos de protección y las alertas de previsión para mitigar posibles daños.

Capítulo 2

Metodología

2.1. Región de estudio

La provincia de Esmeraldas ocupa un territorio de 14.893 km² en donde habitan alrededor de 643.654 personas, según la proyección demográfica del INEC para 2020 está constituida por 7 cantones.

Cantón	Población	Área (km²)
Atacames	41.526	511
Eloy Alfaro	39.739	4.302
Esmeraldas	189.504	1.351
Muisne	28.474	1.265
Quinindé	122.570	3.855
Rioverde	26.869	1.506
San Lorenzo	42.486	3.106

Cuadro 2.1: cantones en Esmeraldas.

Las actividades principales en la provincia de Esmeraldas son el comercio, la ganadería, la industria, la agricultura; y gran parte de su economía depende de la exportación de camarón y banano.

Al estar limitado al oeste y norte con el océano Pacífico, es importante tener control acerca de las actividades oceanográficas que se tienen en la región para emitir alertas tempranas y minimizar los daños materiales.

2.2. EL INOCAR

A partir de 1972, el Instituto Oceanográfico de la Armada del Ecuador (INOCAR) fué elevado a la categoría de Instituto Oceanográfico, y dentro de su misión están: "Desarrollar la caracterización hidro-oceanográfica de los espacios marítimos jurisdiccionales y no jurisdiccionales de interés nacional, para contribuir a garantizar la seguridad a la navegación"[18].

Las principales actividades que realiza el INOCAR son:

- informar el cumplimiento de estudios de ciencias del mar, boletines de olas y corrientes.
- Informes de estudios de tsunamis, así como de cartas de inundación y modelamiento.
- Boletines donde se alerta y monitorea un posible tsunami.
- Resultados de simulación de condiciones Océano-Atmosféricas, avisos meteorológicos y de estado de mar.

2.2.1. La marea

Las mareas son movimientos periódicos de ascenso y descenso de grandes masas de agua causadas por la interacción gravitacional que existe entre la tierra y la luna. El océano experimenta dos mareas altas (pleamar) y dos mareas bajas (bajamar) cada día.

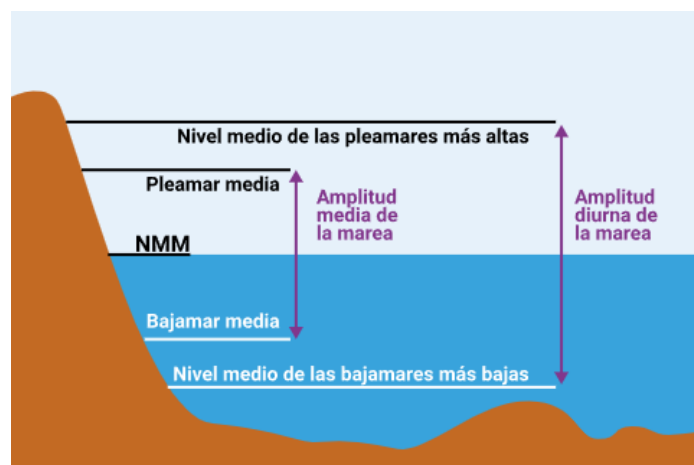


Figura 2.1: Pleamar y bajamar¹

Se identifica marea alta cuando el agua del mar alcanza una altura más alta dentro del ciclo de las mareas, y marea baja se da cuando el agua del mar alcanza su altura más baja [9].

Otros fenómenos atmosféricos como cambios en la presión atmosférica, viento, lluvia provocan variaciones en el nivel del mar. Además, si un viento fuerte sopla en dirección a la costa producirá mareas más altas que las previstas, mientras que un viento en sentido contrario provocará mareas más bajas.

La medición del nivel del mar se utiliza como referencia para la construcción de infraestructura costera y el transporte marino. Las predicciones de mareas que se muestran en la página del INOCAR, se realizan considerando la información de los mareógrafos y el modelo SWAN.

2.2.2. Escala Douglas

La Escala Douglas [12] clasifica los diferentes estados del mar en 9 grados, tiene las características del oleaje, el tamaño y la denominación de las olas tomando como referencia la altura del oleaje.

Grado	Altura de olas (m)	Oleaje	Denominación
0	Sin olas	Sin oleaje	Mar llana o 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 2.2: Escala Douglas

Los cambios del oleaje en la escala Douglas permite poner en alerta a los puertos para tomar medidas de prevención cuando el oleaje alcanza

¹recuperado de [6]

un determinado grado, identificando el estado del mar y el nivel de altura del oleaje, de esta forma evitar pérdidas materiales y humanas [2].

2.3. Tratamiento de datos

Los datos se pueden extraer desde la página del INOCAR en la sección tabla de mareas, seleccionando uno de los 20 puertos disponibles en el mapa se tiene acceso a un documento en formato pdf el cual tiene los datos trimestrales en donde se encuentran las siguientes variables:

- Día y Fecha
- Hora (hhmm)
- Altura (m)

En la provincia de Esmeraldas se tiene información de tres puertos de los cantones: Esmeraldas, Muisne y San Lorenzo. Para el estudio se consideran los datos tomados del cantón Esmeraldas.

El documento pdf descargado no tiene una estructura adecuada para trabajar con los datos directamente, entonces se realizan los siguientes pasos con el fin de obtener una base de datos que sea tratable.

1. Descargar el documento con la información de la altura del oleaje de 2021 disponible en la pagina del INOCAR.
2. Convertir el documento pdf en un archivo de excel desde un software disponible en línea.
3. Estructurar de manera adecuada las variables usando herramientas de transformación de datos en excel, que permitan identificar las variables, los valores de la altura, día, fecha y mes.
4. Cargar la base de datos en R y hacer limpieza de datos.
 - Verificar que la variable mes sea de tipo factor y que existan doce factores correspondientes a los meses en el año.

- Corregir las observaciones que tienen NA en las variables fecha. Los NA representan un espacio en blanco en el cual no se ha completado la información pero tiene los mismos valores del ultimo dato presente. Se puede llenar los espacios usando la función "fill".
- Comprobar que las variables categóricas tengan correcto el tipo de factores.
- Verificar que los valores de la altura sean tipo numérico, para la modelización.

En la base de datos final del cantón Esmeraldas se tiene información desde el 1 de enero de 2021 hasta 31 de diciembre de 2021, se dispone de un total de 1 406 observaciones. Las observaciones cuentan con mediciones tomadas en la mañana, tarde, noche y madrugada, obteniéndose así mediciones de pleamar y bajamar diariamente. Las observaciones de pleamar y bajamar en el mes de enero se pueden ver en la figura 2.2.

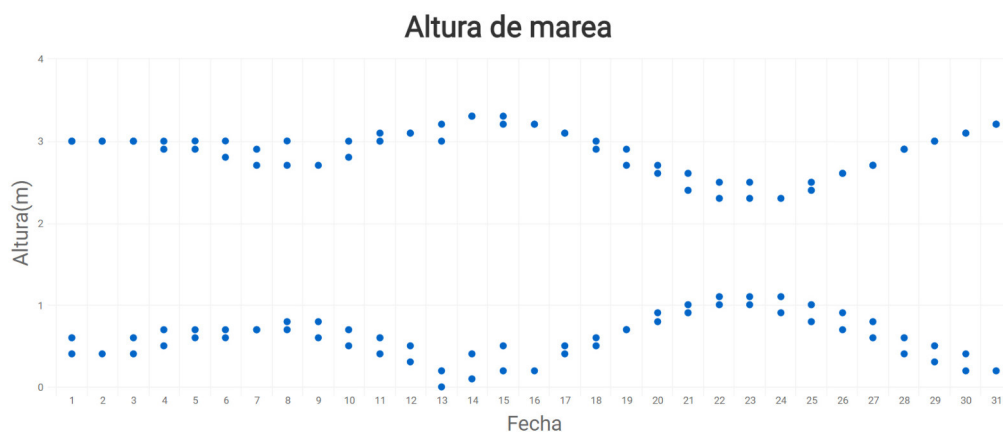


Figura 2.2: Marea enero

Al disponer de dos mediciones de pleamar y dos de bajamar se considera únicamente una medición la cual corresponde a la altura máxima diaria del oleaje ($H_{m\acute{a}x}$). De esta forma se van a considerar 365 datos para el estudio.

Se encontró un dato atípico correspondiente al 31 de marzo de 2021, con una altura de 4.3 [m] que se puede observar en la figura 2.3. El INOCAR indica que se tuvo un aguaje máximo el 30 y 31 de marzo, sin embargo, no se registran fenómenos oceanográficos en esta fecha que

permitan dar motivos de este valor. Se sospecha que es error de medición y por ello se procedió a eliminar esta observación.

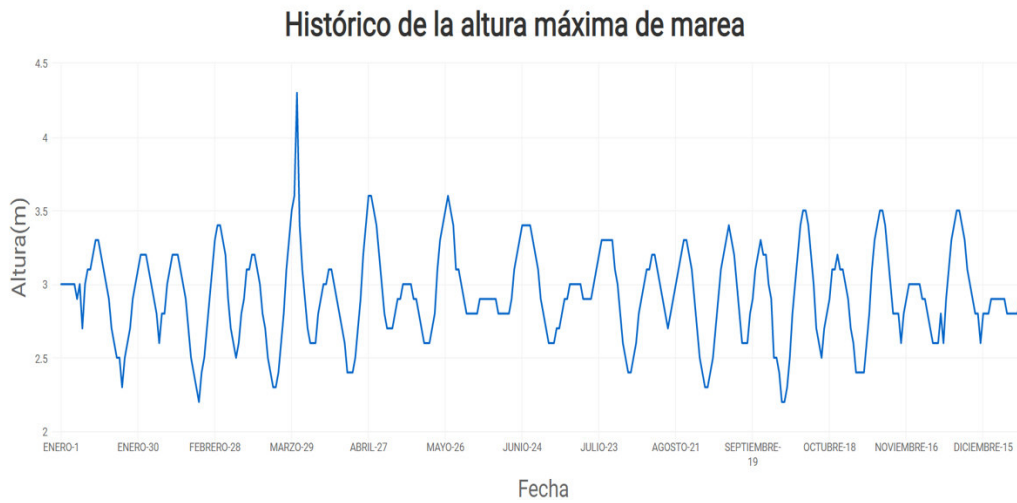


Figura 2.3: $H_{\text{máx}}$ oleaje en Esmeraldas

Debido a la ubicación geográfica en Ecuador se tiene únicamente dos estaciones, las cuales son verano e invierno. Para la modelización se dividió el año en dos períodos como se indica en el cuadro 2.3. No se consideran los datos de diciembre de 2021, pues diciembre es parte del verano del 2022.

Estación	Meses	N°
Invierno	enero - mayo	183
Verano	junio - noviembre	151
Todo	enero - diciembre	365

Cuadro 2.3: Datos por estaciones

En invierno, en la costa ecuatoriana existe presencia de grandes lluvias; en particular en el cantón Esmeraldas, en los meses de abril y mayo se tiene un crecimiento considerable de la altura de las olas, lo cual se puede apreciar en la figura 2.3.

Estadística Descriptiva

En la estación de invierno, se puede visualizar en el diagrama de cajas en la figura 2.4 que no existen datos atípicos. Además en mayo se tienen los valores más altos de la altura máxima del oleaje, mientras que en el mes de febrero se tienen las alturas más bajas.

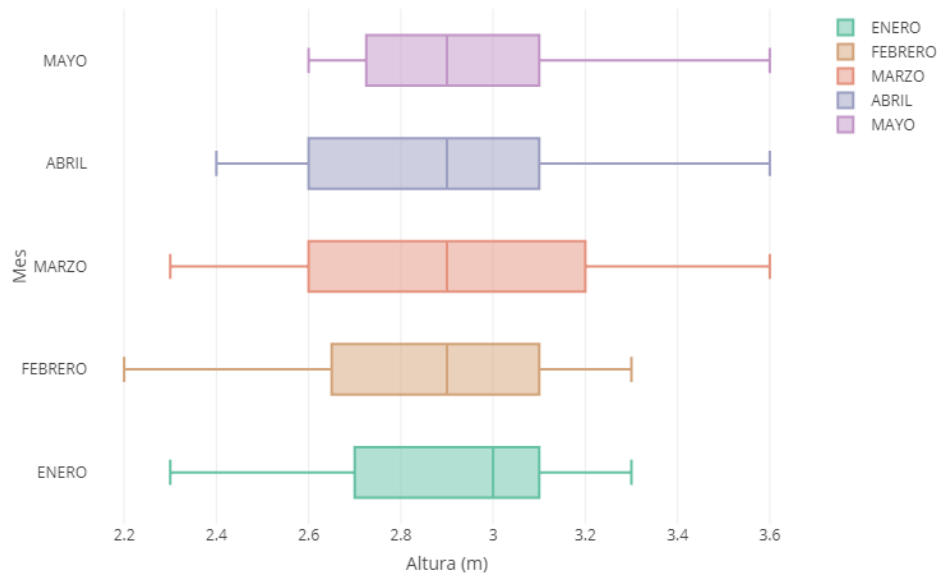


Figura 2.4: Diagrama de cajas invierno

Tomando las medidas de tendencia central, además de los valores mínimos y máximos mensualmente para invierno se tiene los resultados del cuadro 2.4. En donde la mediana para cada mes se encuentra alrededor de 3[m], y existe mayor variabilidad en la altura del mes de marzo.

Mes	Min	Media	Mediana	Max
ENERO	2.3	2.92	3.0	3.3
FEBRERO	2.2	2.86	2.9	3.3
MARZO	2.3	2.91	2.9	3.6
ABRIL	2.4	2.93	2.9	3.6
MAYO	2.6	2.98	2.9	3.6

Cuadro 2.4: Estadística descriptiva en invierno

En invierno se tiene la presencia del fenómeno de niño, es un factor que influye significativamente en la altura máxima ($H_{\text{máx}}$) del oleaje. La

altura de ola significativa corresponde a la media del tercio mayor en la serie de alturas.

Se realizó un análisis análogo para la estación de verano, por lo cual esta se puede visualizar en el anexo [A.2](#).

2.4. Modelización

2.4.1. Distribuciones de probabilidad

En base a la sección [1.4](#), las distribuciones usualmente utilizadas para el moldeamiento del oleaje se analiza las distribuciones candidatas que mejor pueden ajustarse a los datos. Para ello, se utiliza una gráfica de Cullen and Frey que muestra que tipo de distribución se asemejan los datos considerando la asimetría y la curtosis, además para un mejor análisis, se realiza un remuestreo mediante el método de bootstrap de 500 muestras para identificar de manera más eficiente las distribuciones a seleccionar.

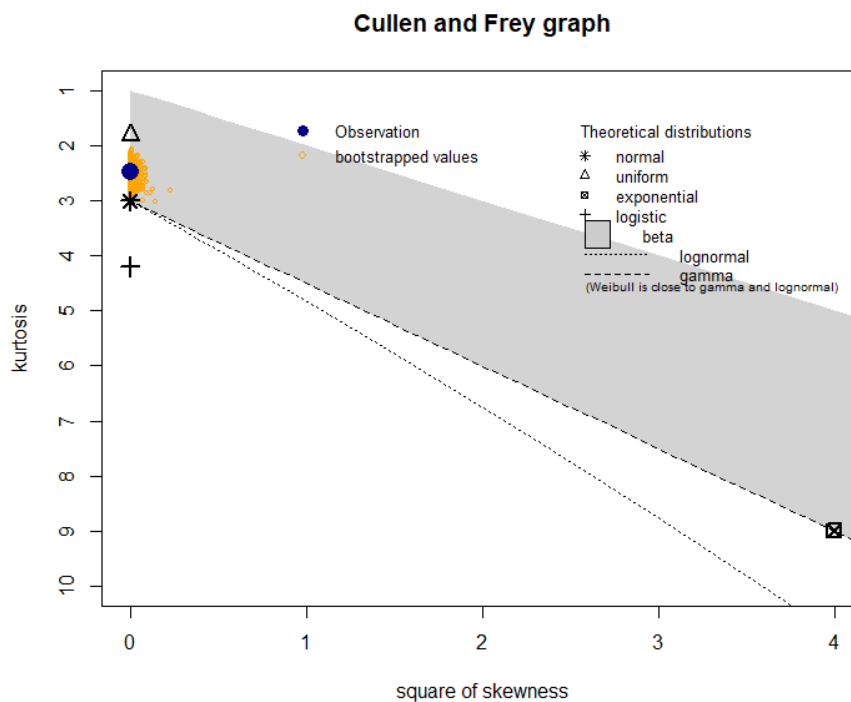


Figura 2.5: Gráfico de Cullen y Frey en invierno

Considerando los datos para invierno, las muestras bootstrap se las muestran de color naranja y los datos observados de color azul en la figura 2.5. Además, se puede observar que las distribuciones Normal, Gamma y Weibull son candidatas para los datos de invierno.

De esta forma para identificar los distribución que mejor se ajustan a los datos se analizarán las distribuciones Normal y Gamma sugeridas por el gráfico de Cullen y Frey, además las distribuciones Gumbel y Weibull que son parte de las distribuciones de valores extremos y al ser estudiadas constantemente en hidrología.

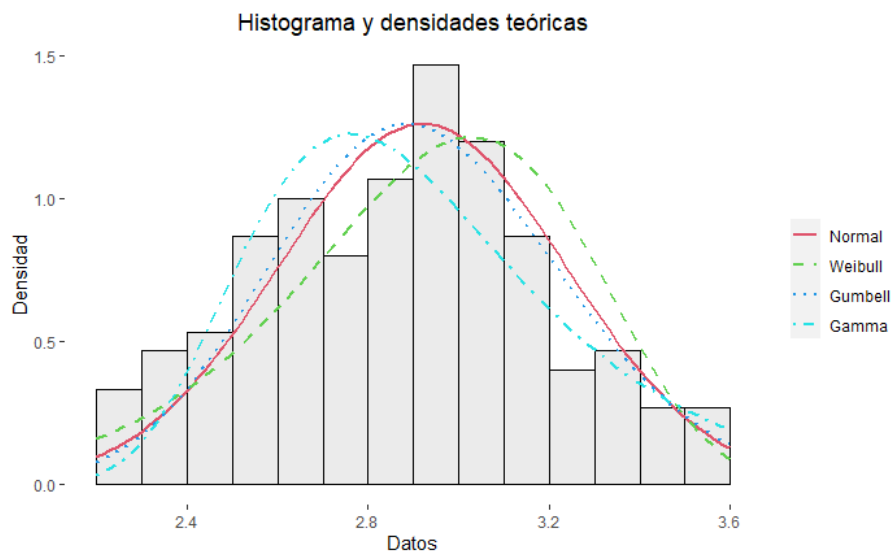


Figura 2.6: Distribuciones invierno

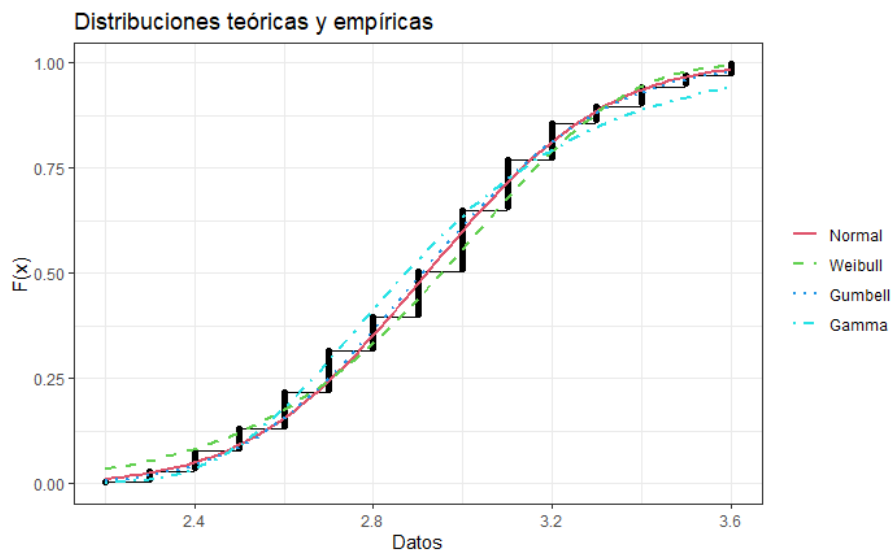


Figura 2.7: Distribuciones acumulada en invierno

En las figuras 2.6 y 2.7 se pueden observar el ajuste de las distribuciones candidatas. La función de distribución de $H_{\text{máx}}$ para la estación de invierno y la comparación con las distribuciones teóricas Normal, Weibull, Gumbel y Gamma.

Para verano se hace el mismo análisis, en las figuras A.3 y A.4 del anexo se puede observar que las cuatro distribuciones mencionadas también son buenas candidatas, se hace un análisis para identificar la mejor distribución que se ajuste a los datos en verano.

2.5. Comparación de los modelos

Para la selección del mejor modelo entre las distribuciones Normal, Weibull, Gamma y Gumbel se hace un contraste de bondad de ajuste para comparar la función de distribución empírica generada por los datos con las funciones de distribución mencionadas anteriormente. Se compararon los estadísticos de prueba de Kolmogorov-Smirnov (KS), Cramer von Mises (CvM) y Anderson-Darling (AD) que se pueden observar en el cuadro 2.5.

	Normal	Weibull	Gamma	Gumbel
KS	0.0940	0.0960	0.1064	0.1323
CvM	0.1545	0.1966	0.1800	0.3771
AD	0.8728	1.3287	0.9805	2.2059

Cuadro 2.5: Pruebas estadísticas invierno.

Al comparar los resultados del cuadro 2.5 se consideran los estadísticos de menor valor que corresponden a las distribuciones Normal y Gamma que tienen un mejor ajuste.

Además, se utiliza el Criterio de Información de Aikake (AIC) y el Criterio de Información Bayesiano (BIC) para comparar las distribuciones, se puede observar en el cuadro 2.6.

	Normal	Weibull	Gamma	Gumbel
AIC	4.4330	92.7556	84.9913	100.8849
BIC	90.4543	98.7768	91.0126	106.9062

Cuadro 2.6: AIC y BIC invierno.

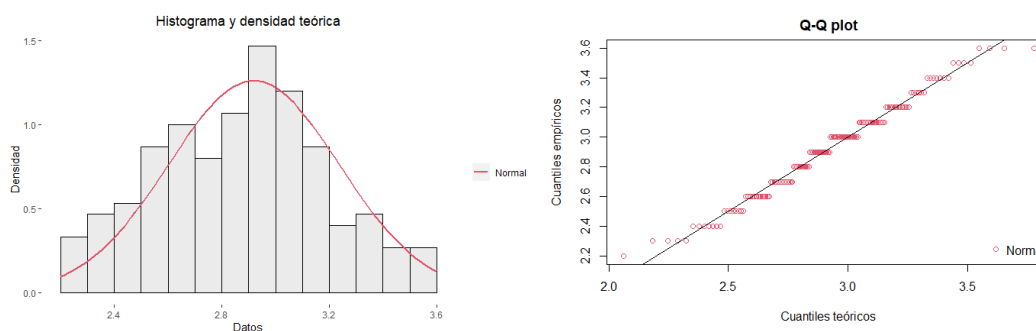
Cuanto más débil sea el criterio mejor se ajustan los datos de invierno a la distribución candidata. En el caso de los datos de invierno los valores menores de AIC y BIC son las distribuciones Normal y Gamma.

Para verano, los resultados se muestran en el anexo en los cuadros A.2 y A.3. Las distribuciones que mejor se ajustan a los datos son las mismas que en el caso de invierno.

De las distribuciones para verano e invierno, se consideran las distribuciones Normal y Gamma, y se van a tomar las distribuciones truncadas considerando las condiciones sobre el conjunto de truncamiento analizado en la sección 1.4.3.

2.5.1. Estimación de los parámetros

Distribución Normal



(a) Distribución e histograma.

(b) QQ-plot distribución Normal

Figura 2.8: Distribución Normal para $H_{\text{máx}}$ en invierno.

Se utilizaron los estadísticos: Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises, en donde los p-valores obtenidos son 0.1441, 0.4312, 0.3764 respectivamente para los datos de invierno. Dado que el p-valor

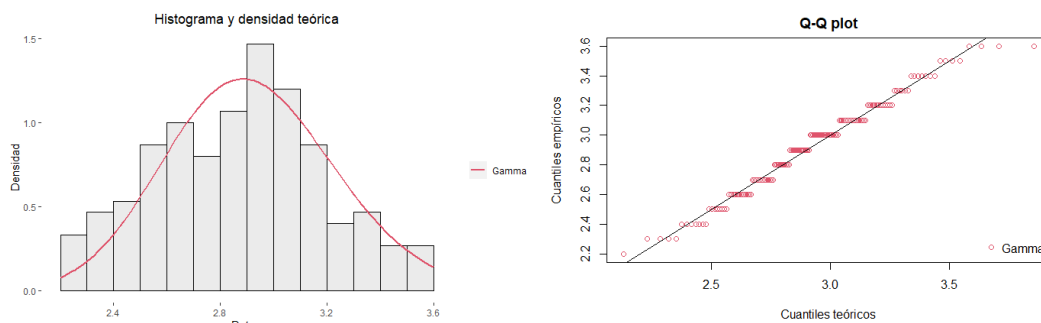
es mayor al nivel de significancia tomando como $\alpha = 0,05$, no se rechaza la hipótesis de que los datos en invierno siguen una distribución Normal.

Para el cálculo de los parámetros se usa el método de máxima verosimilitud para la distribución normal se puede ver en el anexo A.9 los estimadores. Usando el software R se tiene que los parámetros para invierno se obtienen los siguientes estimadores: $\mu = 2.92$; $\sigma = 0.32$.

Para verano se hizo en mismo análisis con las tres pruebas estadísticas, los resultados se muestran en el anexo A.5. De manera similar para los datos para verano siguen una distribución Normal con estimación de los parámetros: $\mu = 2.91$; $\sigma = 0.29$.

Distribución Gamma

De manea similar que en la distribución Normal, para el ajuste de los parámetros se considero el método de máxima verosimilitud.



(a) Distribución e histograma.

(b) QQ-plot distribución Gamma

Figura 2.9: Distribución Gamma para $H_{m\acute{a}x}$ en invierno.

Se utilizaron los estadísticos: Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises, en donde el p-valor es 0.0667, 0.6967, 0.2169 respectivamente para los datos de invierno.

Dado que el p-valor es mayor al nivel de significancia, se tiene que no se rechaza la hipótesis de que los datos en invierno siguen una distribución Gamma.

Para el cálculo de los parámetros se usa el método de máxima verosimilitud para la distribución gamma se puede ver en el anexo A.10 los estimadores. Usando el software R se tiene las estimación de los paráme-

tros para invierno

$$\alpha = 84.15; \theta = 28.83, \text{ donde } \theta = 1/\beta.$$

Para verano se hizo en mismo análisis, los resultados se muestran en el anexo A.6. De esta forma la estimación de los parámetros en verano para la distribución Gamma son: $\alpha = 96.46; \theta = 33.14$

Obtener la estimación de los parámetros para las estaciones de verano e invierno permiten determinar las distribuciones truncadas y determinar el periodo de retorno lo cual se analizará en las siguientes secciones.

2.6. Distribuciones truncadas

Para determinar los parámetros de la distribución normal truncada, se generó una simulación de 2000 datos con los parámetros obtenidos para invierno de esta forma se busca obtener una distribución truncada de esta distribución.

```
simulacion_n <- rNO(2000, mean, sd)
# Distribucion normal truncada en (a,b)
gen.trun(
  par = c(a,b),
  family = "NO",
  name = "_tr",
  type = "both"
)
# simulacion de una distribucion truncada
simulacion_truncada_n <- rNO_tr(2000, mean, sd)
```

La función *gen.trun()*, del paquete *gamlss.tr* permite encontrar una distribución truncada y considera los siguientes argumentos:

- **par**: Considera el intervalo de truncamiento, si el truncamiento es a la derecha o izquierda se considera un solo valor, mientras que *c(a,b)* considera una distribución doblemente truncada.

- **family:** Define la distribución original. Las familias de distribución admitidas por `gamlss()` disponibles en `en` `gamlss.family`, algunas de las distribuciones disponibles son: Norma, Gamma, Gumbel, Weibull, entre otras.
- **name:** Características adicionales que se agregara al nombre de la nueva distribución truncada generada.
- **type:** Tipo de truncamiento tomado, por izquierda ("left"), por derecha ("right"), doblemente truncada ("both")

Para los límites de truncamiento se consideran los casos mencionados en la sección 1.4.3, y se analiza a continuación.

2.6.1. Distribución Normal Truncada

Se conoce que en una distribución Normal $x \in \mathbb{R}$, sin embargo al altura del oleaje es no negativa, y teóricamente no puede tener una altura que se acerque al infinito, entonces se toma la distribución doblemente truncada, en donde el truncamiento por izquierda $a = 0$, y el truncamiento por derecha b , y se analiza en los siguientes casos:

- **Caso I:**

Por la caracterización del oleaje de la escala Douglas y por la estadística descriptiva de la sección 2.3 se conoce que el oleaje en el cantón Esmeraldas se encuentra máximo en grado 5, en donde la altura máxima del oleaje es grueso moderado asociado a esta escala tiene como límite superior $4[m]$, de esta forma el intervalo de truncamiento es $T = [0, 4]$.

- **Caso II:**

Para el punto de truncamiento por derecha se propone tomar el valor máximo de la altura del Oleaje, entonces el intervalo de truncamiento esta en

$$T = [0, \text{máx}(\text{Datos})]$$

■ **Caso III:**

Al identificar los valores máximo y mínimo de la altura máxima del oleaje, se descartan los valores fuera de este rango, entonces el intervalo de truncamiento es comando como:

$$T = [\text{mín}(\text{Datos}), \text{máx}(\text{Datos})]$$

En los datos de invierno la altura máxima es 3,6, entonces el intervalo de truncamiento es tomado $T = [0; 3,6]$

De manera similar en los datos de verano la altura máxima es 3,5, entonces el intervalo de truncamiento es tomado $T = [0; 3,5]$.

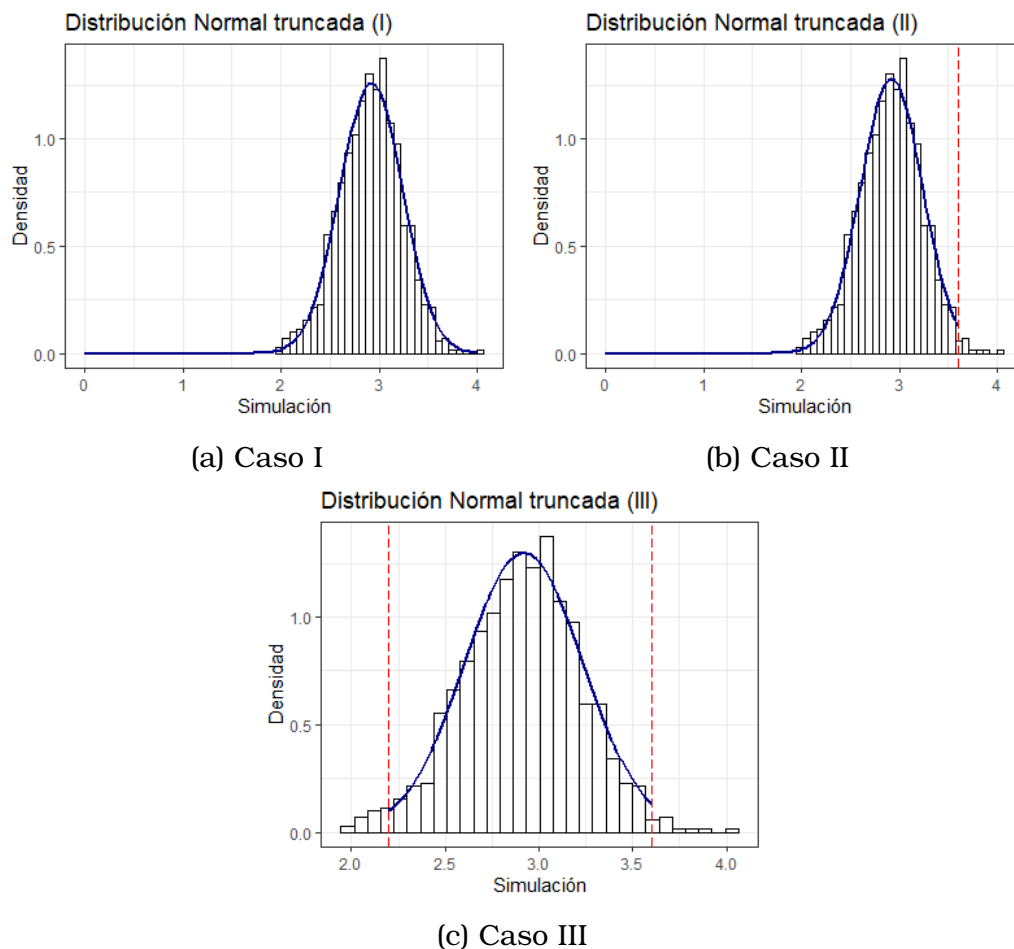


Figura 2.10: Distribuciones Normales truncadas en invierno

Para verificar si las distribuciones Normales truncadas se ajustan a los datos de la altura máxima de oleaje en invierno se realiza el test de

Kolmogorov Smirnov, se muestra en la siguiente tabla los resultados en el cuadro 2.7.

	Caso I	Caso II	Caso III
D	0.0863	0.1143	0.0900
p valor	0.2853	0.0661	0.2413

Cuadro 2.7: Pruebas KS para distribuciones truncadas invierno

Comprobando el nivel de significancia, se tiene que en los tres casos planteados siguen una distribución normal truncada. El caso I; es decir en el intervalo $[0,4]$, presenta mejores resultados para la distribución truncada.

De manera similar, tomando los datos para verano se tiene

	Caso I	Caso II	Caso III
D	0.0851	0.0895	0.0865
p valor	0.2130	0.1676	0.1973

Cuadro 2.8: Pruebas KS para distribuciones truncadas verano

En verano se tienen resultados similares, al considerar la distribución doblemente truncada (caso I), es decir en el intervalo $[0,4]$. Los gráficos de las distribuciones truncadas para los tres casos planteados anteriormente se visualizan en los anexos A.7.

La distribución que modela mejor los datos es la distribución Normal truncada tomando los intervalos de truncamiento del Caso I, $T = [0, 4]$ y por la ecuación 1.2, se tiene

$$f(x) = \frac{\phi_{\mu,\sigma^2}(x)}{\Phi_{\mu,\sigma^2}(4) - \Phi_{\mu,\sigma^2}(0)}, \quad x \in [0, 4]. \quad (2.1)$$

Para invierno, se tiene que la altura máxima del oleaje sigue una distribución Normal $N(\mu = 2,92, \sigma = 0,32)$ truncada en $T = [0; 4]$

Para verano, que tiene que la altura máxima del oleaje sigue una distribución Normal $N(\mu = 2,91, \sigma = 0,29)$ truncada en $T = [0; 4]$

2.6.2. Distribución Gamma truncada

Se conoce que en una distribución Gamma $x \geq 0$, además teóricamente no puede tener una altura que se acerque al infinito, entonces se toma la distribución truncada por derecha en b , y de manera similar a la distribución normal truncada se analiza en los siguientes casos:

- **Caso I:**

Truncando la distribución por derecha, se conoce que el oleaje en el cantón Esmeraldas se encuentra máximo en grado 5, y por la escala Douglas la altura máxima del oleaje es grueso moderado asociado a esta escala tiene como limite superior $4[m]$, de esta forma se toma como 4 el punto de truncamiento por derecha y el intervalo de truncamiento es $T = [0, 4]$.

- **Caso II:** Una alternativa para el punto de truncamiento por derecha se propone tomar el valor máximo de la altura del Oleaje, entonces el intervalo de truncamiento esta en

$$T = [0, \text{máx}(\text{Datos})]$$

- **Caso III:**

Al identificar los valores máximo y mínimo de la altura máxima del oleaje, se descartan los valores fuera de este rango, entonces el intervalo de truncamiento es comando como:

$$T = [\text{mín}(\text{Datos}), \text{máx}(\text{Datos})]$$

En los datos de invierno la altura máxima es 3,6, entonces el intervalo de truncamiento es tomado $T = [0; 3,6]$

En los datos de invierno la altura máxima es 3,6, entonces el intervalo de truncamiento es tomado $T = [0; 3,6]$

De manera similar en los datos de verano la altura máxima es 3,5, entonces el intervalo de truncamiento es tomado $T = [0; 3,5]$.

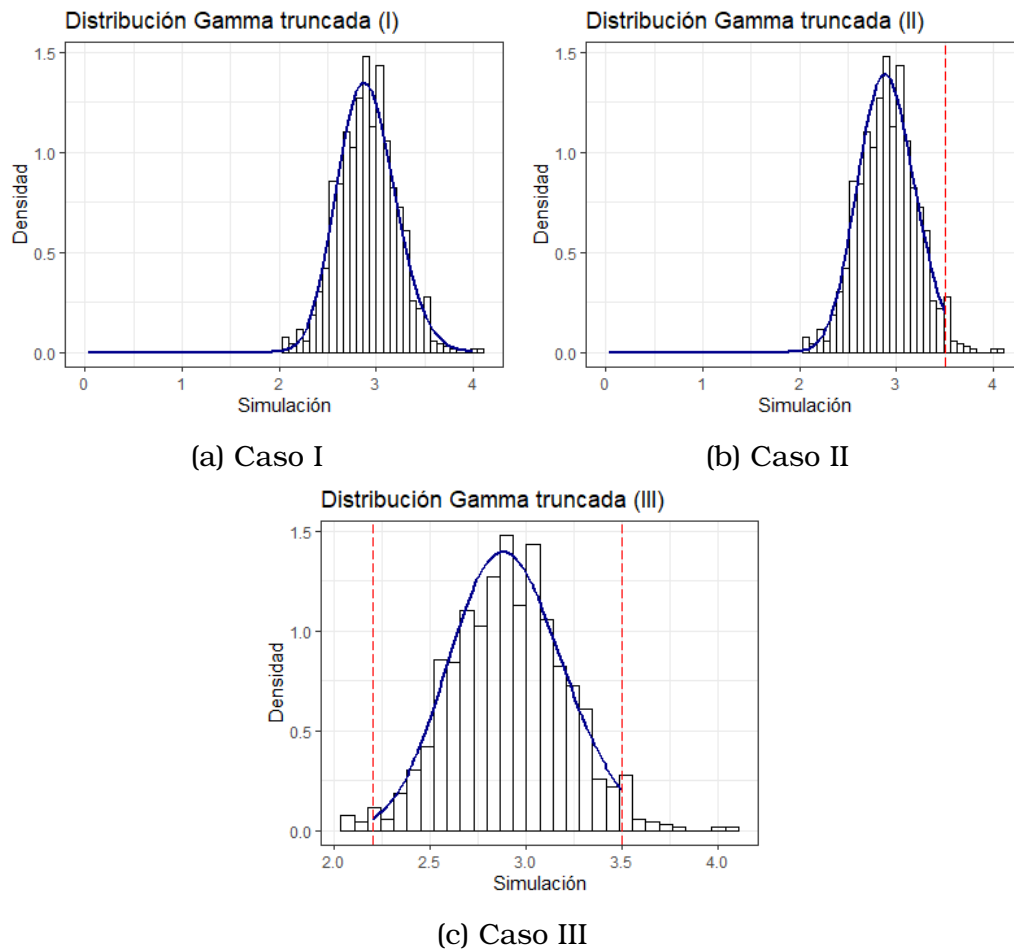


Figura 2.11: Distribuciones Gamma truncadas en invierno

Para verificar si las distribuciones Gamma truncadas se ajustan a los datos de la altura máxima de oleaje en invierno se realiza el test de Kolmogorov Smirnov, se muestra en la siguiente tabla los resultados

	Caso I	Caso II	Caso III
D	0.0853	0.0973	0.1353
p valor	0.2983	0.1688	0.0168

Cuadro 2.9: Pruebas KS para distribuciones truncadas invierno

Comprobando el nivel de significancia, se tiene que en los dos casos planteados siguen una distribución Gamma truncada. El caso I; es decir

truncada por derecha en el intervalo $[0,4]$, y el caso II, es decir en el intervalo $[0;3,6]$ presenta mejores resultados para la distribución truncada, se toma la distribución truncada en el Caso I.

	Caso I	Caso II	Caso III
D	0.0867	0.0922	0.1247
p valor	0.1952	0.1444	0.0163

Cuadro 2.10: Pruebas KS para distribuciones truncadas verano

En verano se tienen resultados similares, es decir al considerar la distribución Gamma truncada por derecha (Caso I) para los datos de la altura máxima.

Los gráficos para verano de las distribuciones Gama truncadas para los tres casos planteados anteriormente se visualizan en los anexos [A.8](#).

La distribución que modela los datos es la distribución Gamma truncada tomando los intervalos de truncamiento del Caso I, $T = [0, 4]$ y por la ecuación [1.2](#), en donde $f(x) = \Gamma(\alpha, \beta)$.

Para invierno, se tiene que la altura máxima del oleaje sigue una distribución Gamma, con parámetro de forma $\alpha = 84.15$ y tasa $\theta = 28.83$.

Para verano, que tiene que la altura máxima del oleaje sigue una distribución Gamma, con parámetro de forma $\alpha = 96.46$ y tasa $\theta = 33.14$.

Se considera el conjunto de truncamiento es $T = [0; 4]$.

2.7. Período de retorno

Al tomar los datos de invierno, para la altura máxima diaria de las olas, el período de retorno se establece como T . La altura máxima diaria de las olas de la zona marítima es una sucesión de variables aleatorias independientes e idénticamente distribuidas X_1, X_2, \dots, X_n . Sea F la distribución de X_1 encontrada en la sección anterior, entonces

$$T = \frac{1}{P(X > x_T)} = \frac{1}{1 - P(X \leq x_T)} = 1 - \frac{1}{T}$$

donde x_T es la altura del oleaje al tiempo T y representa el nivel de retorno por primera vez.

Sea N es el tiempo de parada definido por $N = \inf\{n \geq 1 : X_n > x_T\}$ que indica el momento cuando el período de retorno T y la primera vez que se supera el nivel de retorno correspondiente.

En el Teorema 2 en la sección 1.4.5, se demostró que el tiempo de parada sigue una distribución geométrica del parámetro p , entonces

$$p = P(X_1 > x_T) = 1 - F(x_T)$$

Si se conoce la función de distribución de la altura de ola extrema diaria, entonces se puede obtener la distribución del tiempo de parada.

En la sección 2.6.1 se encontró los parámetros estimados para la distribución Normal truncada que son: $\mu = 2,91$ y $\sigma = 0,29$ para la estación de verano y por la ecuación 2.1 se determina el período de retorno de la altura máxima del oleaje diario, los valores de encuentran en grado cinco en referencia a la escala Douglas.

Los valores del periodo de retorno, la probabilidad de ocurrencia y el nivel de retorno del oleaje se muestra en el cuadro 2.11.

Nivel de retorno (m)	Período de retorno (días)	Probabilidad
2.2	1.0078	0.9923
2.4	1.0427	0.9590
2.6	1.1696	0.8549
2.8	1.5465	0.6466
3.0	2.6320	0.3799
3.2	6.1848	0.1617
3.4	21.0677	0.0474
3.6	107.8857	0.0093

Cuadro 2.11: Período de retorno para verano

En la sección 2.6.1 se encontró los parámetros estimados para la distribución Normal truncada que son: $\mu = 2,91$ y $\sigma = 0,29$ para la estación de verano y por la ecuación 2.1 se determina el período de retorno de la altura máxima del oleaje diario, los valores de encuentran en grado cinco en referencia a la escala Douglas

De manera similar para invierno, En la sección 2.6.2 se encontró los parámetros estimados para la distribución Gamma truncada que son: $\alpha = 84.15$ y tasa $\theta = 28.83$. y por la ecuación 1.2 y la 1.1 se determina el período de retorno de la altura máxima del oleaje diario.

Los valores del periodo de retorno, la probabilidad de ocurrencia y el nivel de retorno del oleaje se muestra en el cuadro 2.12.

Nivel de retorno (m)	Período de retorno (días)	Probabilidad
2.2	1.0048	0.9952
2.4	1.0373	0.9641
2.6	1.1713	0.8537
2.8	1.5770	0.6341
3.0	2.7086	0.3692
3.2	6.1320	0.1631
3.4	18.4349	0.0542
3.6	73.9320	0.0135

Cuadro 2.12: Período de retorno para invierno

En los cuadros 2.12 y 2.11 se observan las deferencias del periodo de retorno al considerar las distribuciones truncadas.

2.8. Aplicativo Shiny

Se desarrolló un aplicativo en Shiny, el cual determina la función de distribución que mejor se ajusta a los datos de las alturas diarias de pleamar y bajamar disponibles del cantón Esmeraldas desde el 1 de enero de 2021 al 31 de diciembre de 2021. En el aplicativo está disponible la opción de cargar nuevos datos.

Para realizar el mismo análisis deben estar presentes las variables: mes, día, fecha, hora y altura. Se descarga la tabla de mareas de la página del INOCAR.

El aplicativo hace una limpieza de datos y genera una nueva base en el cual se incluye la altura diaria máxima del oleaje para cada estación: invierno y verano. Presenta estadísticas descriptivas de los datos, el ajuste y las pruebas de hipótesis de las distribuciones: Normal, Gamma, Gumbel, Weibull, y las distribuciones truncadas para cada estación. La estructura del aplicativo desarrollado se observan en las figuras 2.12 y 2.13. ²

²La versión final de aplicativo estará disponible en marzo 2022.

Modelización de Marejadas - TIC

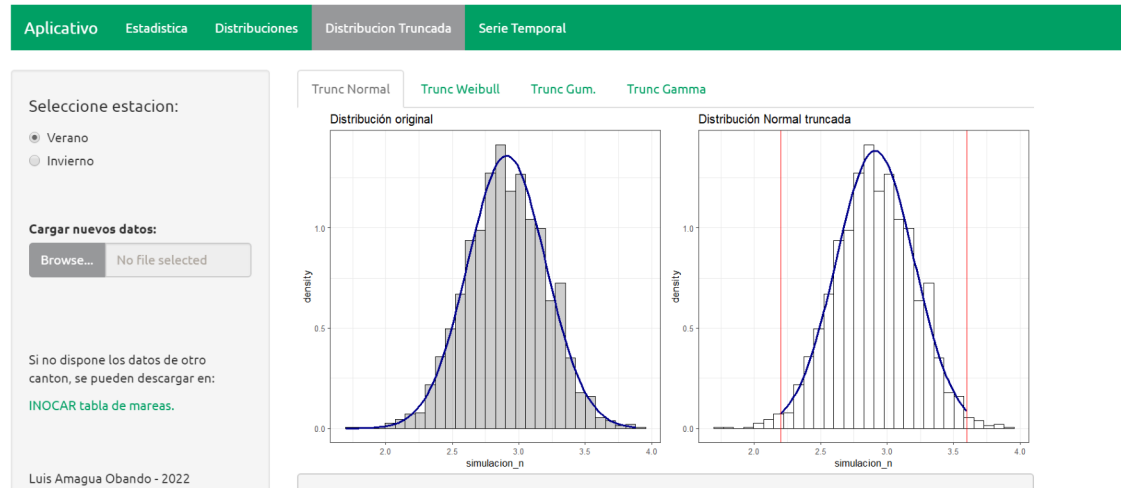


Figura 2.12: Aplicativo Shiny: Distribuciones truncadas

Modelización de Marejadas - TIC

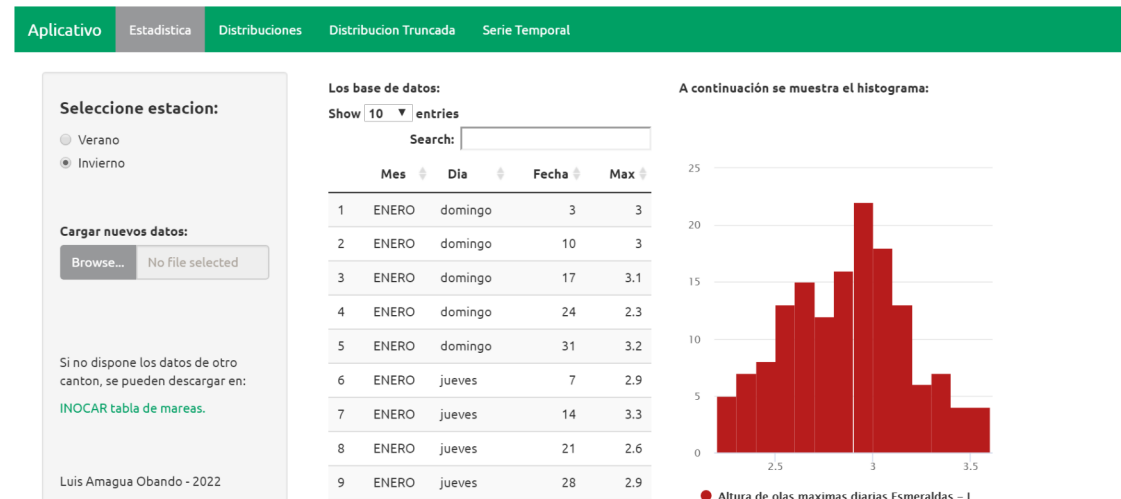


Figura 2.13: Aplicativo Shiny: Estadística Básica

Capítulo 3

Resultados, conclusiones y recomendaciones

En este capítulo se muestran los principales resultados obtenidos de la modelización del oleaje en el cantón Esmeraldas. La diferencia entre distribuciones truncadas y las distribuciones originales, así como los resultados del período de retorno a partir de distribuciones truncadas en las estaciones de verano e invierno.

3.1. Resultados y Conclusiones

3.1.1. Para Invierno

Para determinar la distribución que mejor se ajuste a los datos es necesario hacer pruebas de hipótesis adicionales al test de Komogorvo-Smirnov, se observa en el gráfico 2.5 que las distribuciones candidatas son las distribuciones Normal y Weibull, y al realizar las pruebas de hipótesis de AD, CvM, BIC y AIC se tiene certeza de que las distribuciones mencionadas pueden modelar a los datos de invierno.

Por el principio de parsimonia, para los datos es posible determinar distribuciones más simples que las distribuciones de valores extremos como se planeaba inicialmente. En este estudio se pudo determinar que las distribuciones Normal y Gamma se ajustaron al conjunto de datos en invierno.

La escala Douglas permite determinar los puntos de truncamiento y muestran un mejor ajuste respecto al tomar como punto de truncamiento el valor máximo del conjunto de datos. Para los datos de invierno y verano se tomaron la distribución truncada en $T = (0, 4)$.

Se compara el p-valor y el valor del estadístico KS de las cuatro distribuciones. Como en el cantón Esmeraldas no presenta eventos oceanográficos de gran magnitud, la distribución del oleaje se ajusta a una distribución Normal y una distribución Gamma. Los resultados para la estación de invierno, se muestran en el cuadro 3.1.

Distribución	p valor	D
Normal	0.1441	0.0936
Weibull	0.1261	0.0959
Gamma	0.0668	0.1064
Gumbel	0.0253	0.1323

Cuadro 3.1: KS para distribuciones en invierno

Debido a que el p-valor es mayor al nivel de significancia no se rechaza la hipótesis de que los datos del oleaje sigan distribución Normal, Weibull o Gamma, además por el cuadro 2.5 y el cuadro 2.6 se toman las distribuciones Normal y Gamma como las distribuciones que mejor se ajustan a los datos de invierno.

Se analiza el p-valor para la comparación de las distribuciones truncadas para invierno y verano en el cuadro 3.2.

Distribución	invierno	verano
Normal Truncada	0.2853	0.2130
Gamma Truncada	0.2983	0.1952

Cuadro 3.2: p-valor de distribuciones truncadas

Al considerar un nivel de confianza de 95%, la distribución Gamma truncada se ajusta mejor a los datos de invierno.

En la sección 2.5.1 se determinó la estimación de los parámetros. La distribución que mejor se ajusta a los datos para la estación de invierno

es la distribución Gamma truncada, en donde los parámetros de la distribución son: $\alpha = 84,15$; $\theta = 28,83$, obtenidos de la sección 2.6.2.

Al comparar los resultados del p valor del cuadro 3.2 de las distribuciones truncadas y los cuadros 3.1 y 3.3 de las distribuciones originales, se tiene un incremento en el p valor; es decir, las distribuciones truncadas modelan de mejor forma los datos.

Se puede observar en la Figura 3.1 y en el cuadro 2.12 que cuando el periodo de retorno se incrementa, se tiene que el nivel de retorno se incrementa gradualmente, además se espera que 74 días la altura del oleaje alcance nuevamente una altura de 3.6 m.

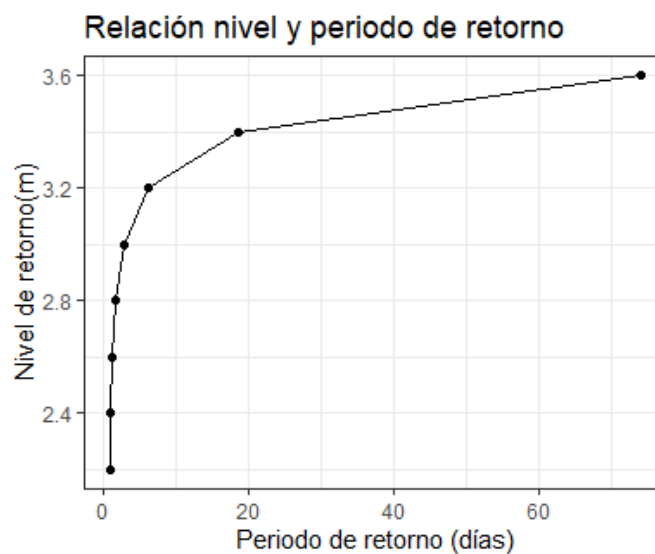


Figura 3.1: Altura vs periodo de retorno

En la figura 3.2, el parámetro de la distribución del tiempo de parada se reducen gradualmente, lo que indica que el nivel de retorno está inversamente relacionado con el parámetro de la distribución del tiempo de parada.

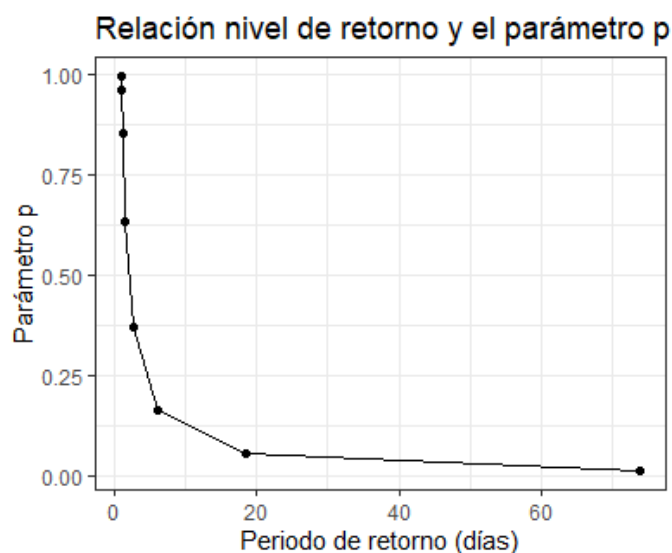


Figura 3.2: Altura vs periodo de retorno

3.1.2. Para Verano

Al considerar las distribuciones en la estación de verano se observa en la figura A.2 que la distribuciones Normal y Gamma son las mejores candidatas, y al comparar las pruebas de hipótesis de AD, CvM, BIC y AIC se determina que la distribución candidata es la distribuciones Normal, se puede observar en los cuadros A.2 y A.3.

Por el principio de parsimonia, se pudo encontrar que las distribución Normal se ajustó al conjunto de datos en verano.

Se compara el p-valor y el valor del estadístico KS de las cuatro distribuciones, la distribución del oleaje se ajusta a una distribución Normal. Los resultados para la estación de verano se muestran en el cuadro 3.3.

Distribución	p valor	D
Normal	0.0937	0.0911
Wiebull	0.0116	0.1186
Gamma	0.0399	0.1034
Gumbel	0.0012	0.1534

Cuadro 3.3: KS para distribuciones en verano

Debido a que el p-valor es mayor al nivel de significancia no se rechaza

la hipótesis de que los datos del oleaje sigan distribución Normal, además por el cuadro A.2 y el cuadro A.3 se toman las distribución Normal como las distribución que mejor se ajustan a los datos de verano.

Se analiza el p-valor para la comparación de las distribuciones truncadas para verano en el cuadro 3.2 .

Al considerar un nivel de confianza de 95 %, la distribución Normal truncada se ajusta mejor a los datos de verano.

La distribución que mejor se ajusta a los datos para la estación de verano es la distribución Normal truncada, en donde los parámetros de la distribución son: $\mu = 2,91$; $\sigma = 0,29$, obtenidos de la sección 2.6.1

Se puede observar en la Figura 3.3 y en el cuadro 2.11 que cuando el periodo de retorno se incrementa, se tiene que el nivel de retorno se incrementa gradualmente, además se espera que 108 días la altura del oleaje alcance nuevamente una altura de 3.6 m.

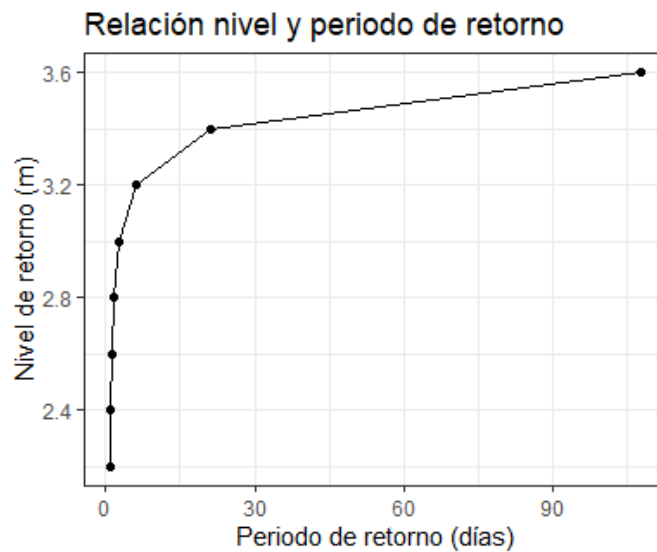


Figura 3.3: Altura vs periodo de retorno en verano.

En la figura 3.4, el parámetro de la distribución del tiempo de parada se reducen gradualmente, lo que indica que el nivel de retorno está inversamente relacionado con el parámetro de la distribución del tiempo de parada.

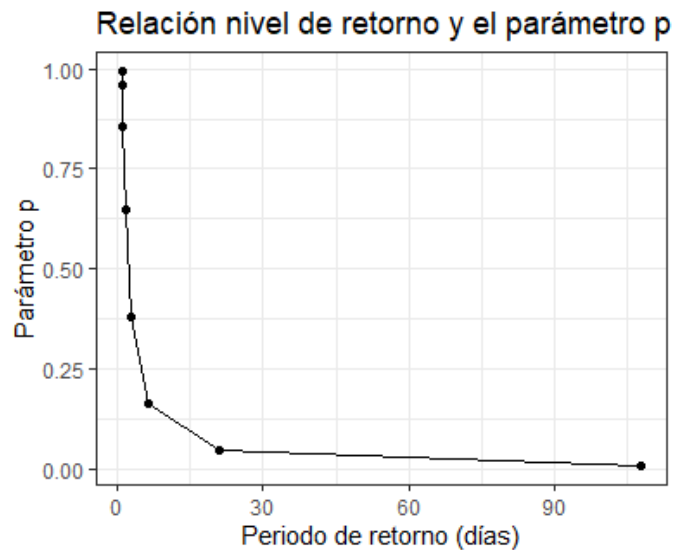


Figura 3.4: Altura vs periodo de retorno en verano.

3.2. Recomendaciones

1. Para ampliar el estudio del oleaje y tener un mejor alcance es necesario tener disponibilidad de datos completos, por ejemplo la dirección del oleaje, velocidad del viento, amplitud de olas, etc.
2. Como trabajo futuro, en la literatura se observó que se pueden hacer comparaciones entre los modelos de las distribuciones planteadas y el fractal autoafin, considerando la regla de Hurst.
3. Como trabajo futuro, existe limitado estudio de predicciones del oleaje mediante el uso de redes neuronales con el modelo LSTM (Long-Short Term Memory), artículos de 2020 efectúan una comparación de rendimientos entre modelos basados en redes neuronales recurrentes, y un modelo de predicción autorregresivo cuyos parámetros se calculan utilizando las ecuaciones de Yule-Walker.

Capítulo A

Anexos

A.1. Teorema tiempo de parada

Sean $\{X_1, X_2, \dots\}$ variables aleatorias independientes e idénticamente distribuidas que representan la altura máxima del oleaje en el espacio de probabilidad (Ω, \mathcal{F}, P) . Si T es el periodo de retorno y $N = \inf\{n \geq 1 : X_n > x_0\}$ entonces el tiempo de parada obedece una distribución geométrica.

Demostración:

$$P(N = k) = P(X_1 \leq x_0, X_2 \leq x_0, \dots, X_{k-1} \leq x_0, X_k > x_0)$$

como $\{X_1, X_2, \dots\}$ son independientes e idénticamente distribuidos

$$P(N = k) = P(X_1 \leq x_0)P(X_2 \leq x_0) \cdots P(X_{k-1} \leq x_0)P(X_k > x_0)$$

si se define $P(X_k > x_0) = p$, entonces se tiene

$$P(N = k) = (1 - p)(1 - p) \cdots (1 - p)p$$

$$P(N = k) = (1 - p)^k p$$

Es decir, se tiene que el tiempo de parada N sigue una distribución geométrica de parámetro p .

Ahora, supongamos que $X \sim Geom(p)$ entonces la función de probabilidad es

$$P(X = x) = p(1 - p)^x$$

Entonces

$$E(X) = 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + \dots$$

$$E(x) = p + 2p(1 - p) + 3p(1 - p)^2 + \dots$$

$$(1 - p)E(X) = p(1 - p) + 2p(1 - p)^2 + 3p(1 - p)^3 + \dots$$

$$(p - 1)E(X) = -p(1 - p) - 2p(1 - p)^2 - 3p(1 - p)^3 + \dots$$

podemos sumar $E(X)$ y $(p - 1)E(X)$, entonces

$$E(X) + (p - 1)E(x) = p + p(1 - p) + p(1 - p)^2 + \dots$$

$$pE(X) = p + p(1 - p) + p(1 - p)^2 + \dots$$

$$E(X) = 1 + (1 - p) + (1 - p)^2 + \dots$$

Podemos reescribir como una serie geométrica, y se tiene

$$E(x) = \sum_{k=0}^{\infty} (1 - p)^k$$

como $p \in (0, 1)$ por la convergencia de la serie, entonces

$$E(x) = \frac{1}{1 - (1 - p)} = \frac{1}{p}$$

Sabemos además que

$$P(X_1 > x_0) = \frac{1}{T}$$

entonces $p = \frac{1}{T}$

Finalmente se tiene

$$EN = \frac{1}{p}$$

A.2. Estadística Descriptiva

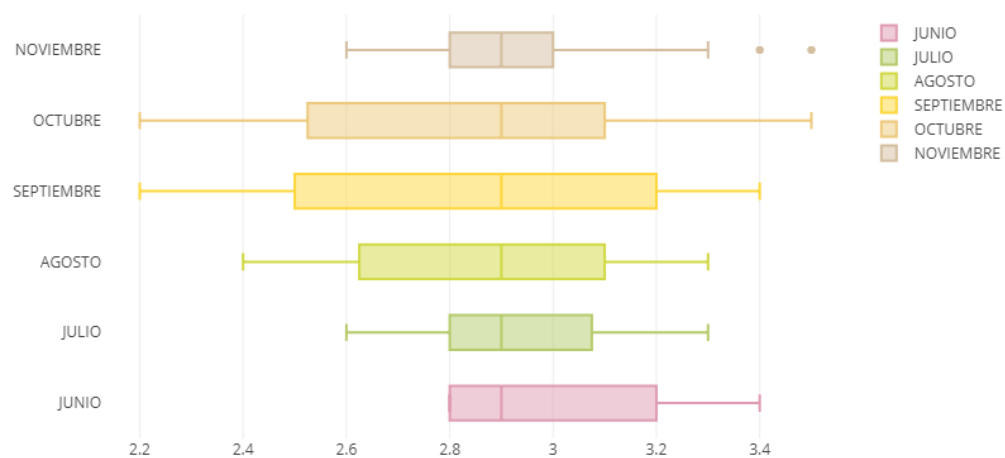


Figura A.1: Diagrama de cajas verano

Mes	Min	Media	Mediana	Max
JUNIO	2.8	3.00	2.9	3.4
JULIO	2.6	2.95	2.9	3.3
AGOSTO	2.4	2.88	2.9	3.3
SEPTIEMBRE	2.2	2.85	2.9	3.4
OCTUBRE	2.2	2.86	2.9	3.5
NOVIEMBRE	2.6	2.93	2.9	3.5

Cuadro A.1: Estadística descriptiva en verano

A.3. Distribuciones para verano

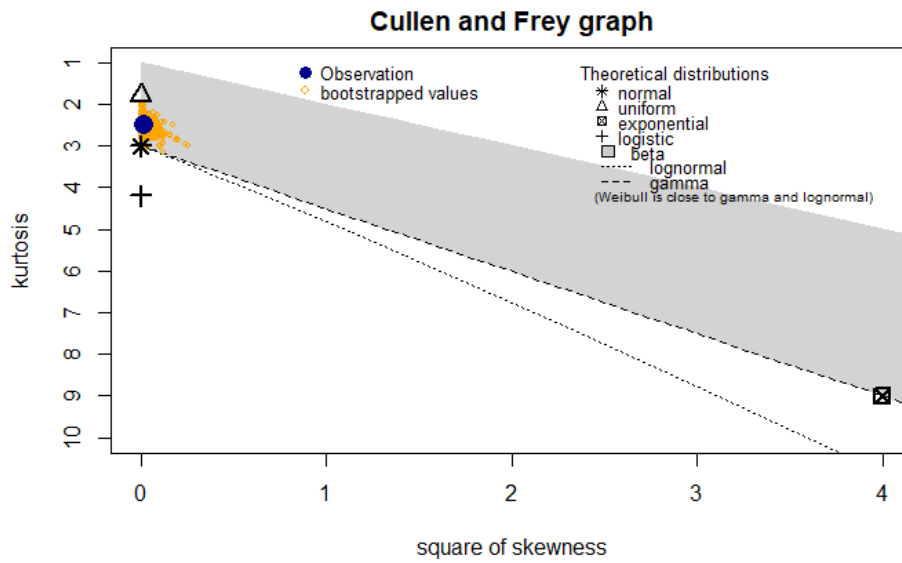


Figura A.2: Gráfico de Cullen y Frey en verano

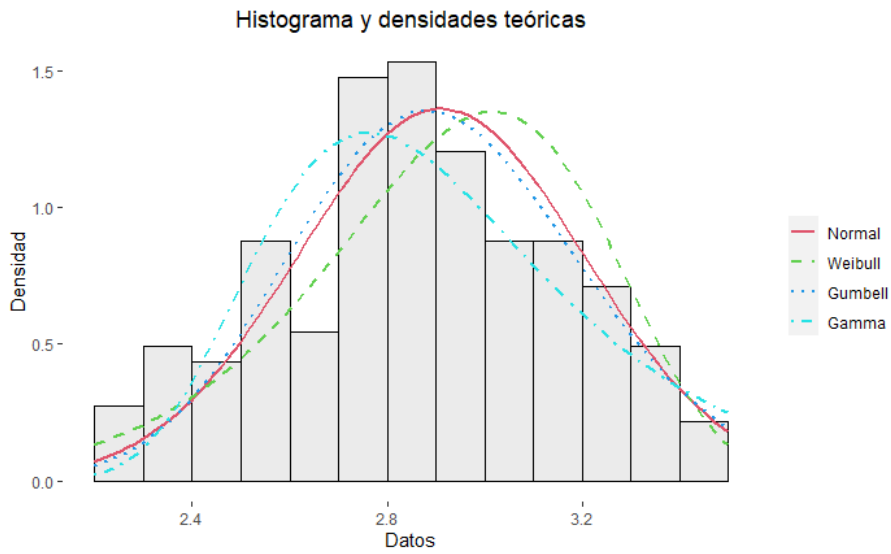


Figura A.3: Distribuciones verano

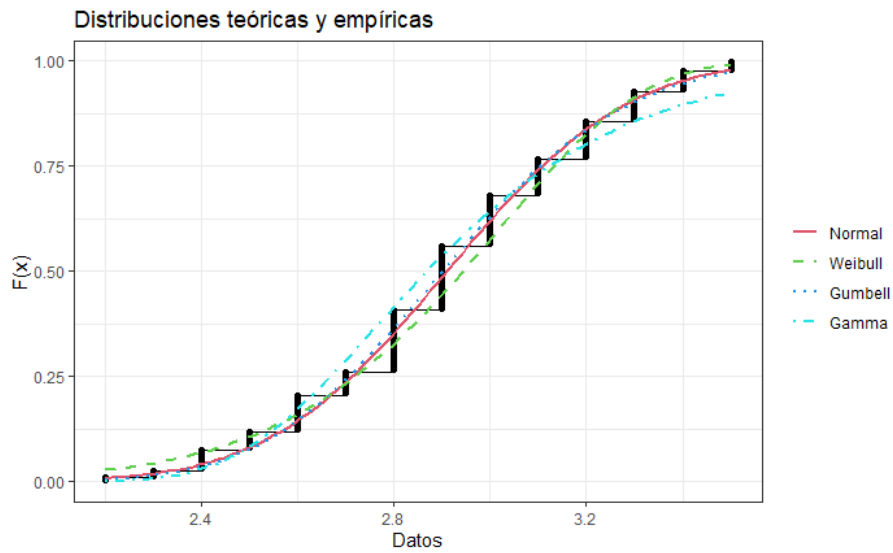


Figura A.4: Distribuciones acumulada en verano

A.4. Pruebas de hipótesis verano

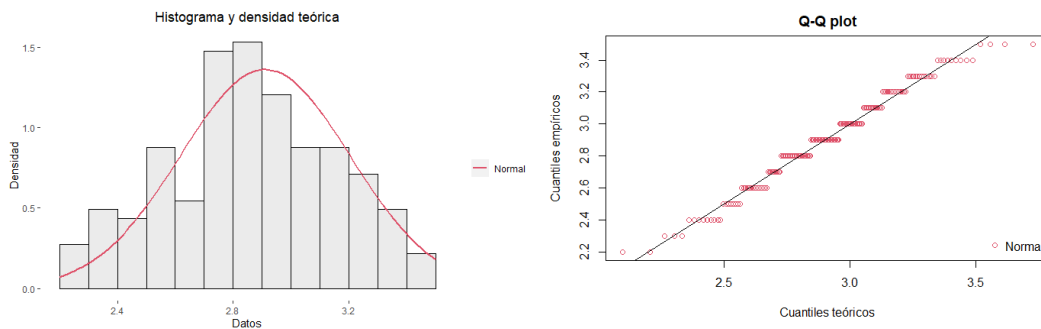
	Normal	Weibull	Gamma	Gumbel
KS	0.0911	0.1186	0.1034	0.1534
CvM	0.2070	0.2974	0.2318	0.5329
AD	1.2188	1.6855	1.3848	3.2968

Cuadro A.2: Pruebas estadísticas verano.

	Normal	Weibull	Gamma	Gumbel
AIC	74.5005	79.1315	76.90438	103.6772
BIC	80.9195	85.5505	83.32335	110.0961

Cuadro A.3: AIC y BIC verano.

A.5. Distribución Normal verano

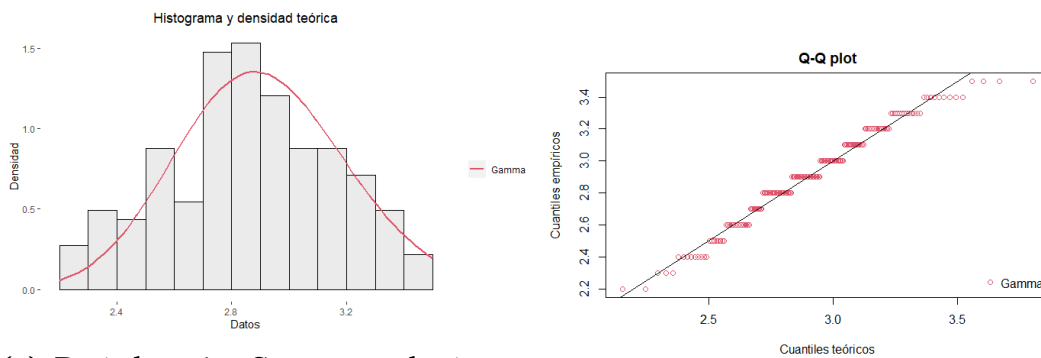


(a) Distribución Normal e histograma. (b) Q-Q-plot distribución Normal

Figura A.5: Distribuciones para $H_{\text{máx}}$ en verano.

Se utilizaron los estadísticos: Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises, en donde los p-valores obtenidos son 0.0937, 0.2604, 0.2237 respectivamente para los datos de verano. Dado que el p-valor es mayor al nivel de significancia, tomando como $\alpha = 0,05$ se que tiene los datos en verano siguen una distribución normal.

A.6. Distribución Gamma verano



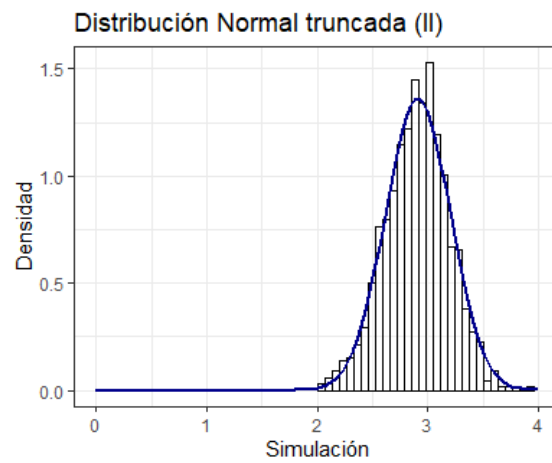
(a) Distribución Gamma e histograma.

(b) Q-Q-plot distribución Gamma

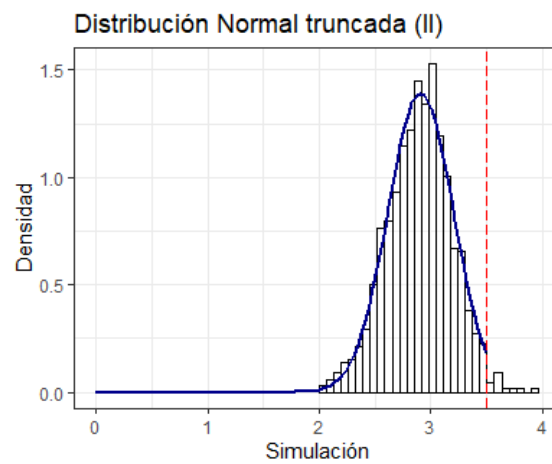
Figura A.6: Distribución Gamma para $H_{\text{máx}}$ en verano.

Se utilizaron los estadísticos: Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises, en donde el p-valor es 0.0667, 0.7813, 0.2237 respectivamente para los datos de verano.

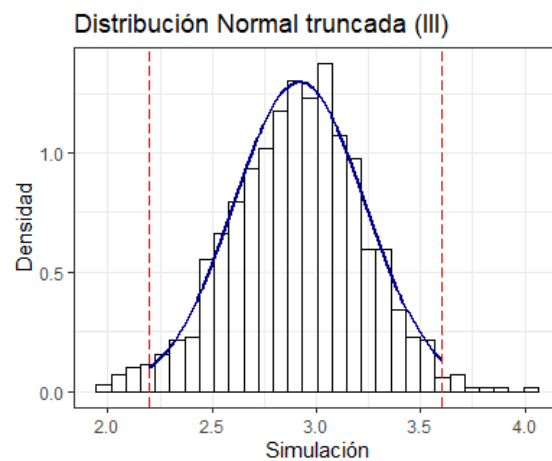
A.7. Distribución Normal truncadas verano



(a) Caso I



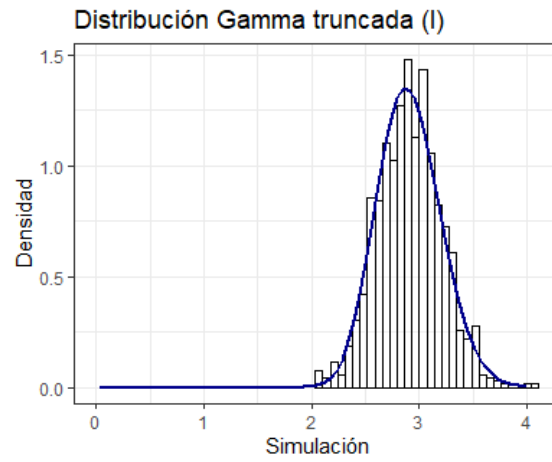
(b) Caso II



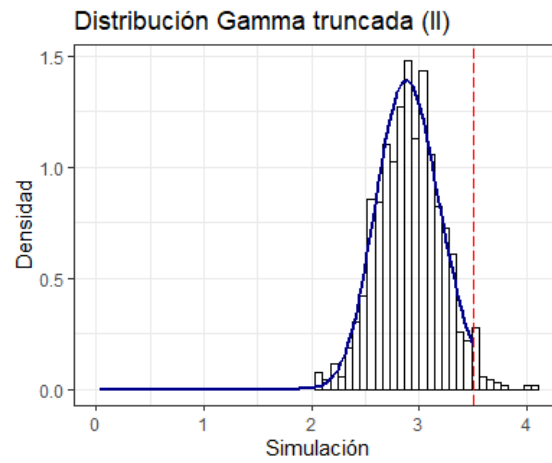
(c) Caso III

Figura A.7: Distribuciones Normales truncadas en verano

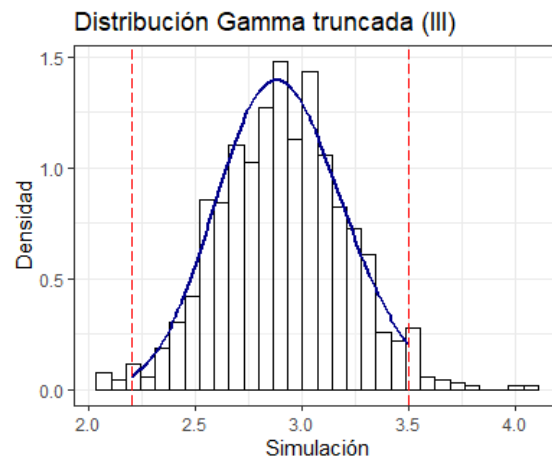
A.8. Distribución Gamma truncada



(a) Caso I



(b) Caso II



(c) Caso III

Figura A.8: Distribuciones Gamma truncadas en verano

A.9. EMV de la distribución Normal

Sea $X \sim N(\mu, \sigma^2)$, usando la función de densidad se tiene

$$\mathcal{L}_n(\mu, \sigma^2) = \prod_{i=1}^n f(x_i|\mu, \sigma^2) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \right) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2}$$

tomando el log,

$$\begin{aligned} \log(\mathcal{L}(\mu, \sigma^2)) &= \log \left(\frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2} \right) \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

Para maximizar se deriva la función log-verosimilitud e iguala a cero y se encuentra el máximo de los parámetros

$$\hat{\mu} = \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

A.10. EMV de la distribución Gamma truncada

Sea $X \sim \text{Gamma}(\alpha, \beta)$

$\mathcal{L}_n(\alpha, \beta) = \prod_{i=1}^n f(x_i|\alpha, \beta) = \prod_{i=1}^n \frac{1}{\beta^\alpha \Gamma(\alpha)} x_i^{\alpha-1} e^{-x_i/\beta} = \frac{1}{(\beta^\alpha \Gamma(\alpha))^n} \prod_{i=1}^n x_i^{\alpha-1} e^{-x_i/\beta}$
tomando el log,

$$\log(\mathcal{L}_n(\alpha, \beta)) = -n \log(\Gamma(\alpha)) - n\alpha \log(\beta) + (\alpha - 1) \sum_{i=1}^n \log x_i - \frac{1}{\beta} \sum_{i=1}^n x_i$$

Para maximizar se deriva la función log-verosimilitud e iguala a cero y se resuelven las ecuaciones

$$0 = \frac{\partial}{\partial \alpha} \log(\mathcal{L}_n(\alpha, \beta)) = -n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n \log(\beta) + \sum_{i=1}^n \log(x_i)$$

$$0 = \frac{\partial}{\partial \beta} \log(\mathcal{L}_n(\alpha, \beta)) = \frac{1}{\beta} \left(-n\alpha + \frac{1}{\beta} \sum_{i=1}^n x_i \right)$$

Para el estimador β se tiene

$$\beta = \frac{\bar{x}}{\alpha}$$

Mientras para α no se puede encontrar una función explicita, y se aproxima el valor por métodos numéricos.

Referencias bibliográficas

- [1] H. Aksoy. *Use of Gamma Distribution in Hydrological Analysis*. Turk J Engin Environ Sci, 2000.
- [2] D. Zhao F. Wu S. Liu A. Chia L. Wang. B. Chen, Y. Kou. *Calculations on stopping time and return period*. Springer Nature, 2020.
- [3] A. Clifford Cohen. *Truncated and censored samples, theory and applications*. University of Georgia, 1991.
- [4] Stuart Coles. *An introduction to Statistical Modeling of Extreme Values*. Springer, 2001.
- [5] Centro Nacional de Huracanes de los Estados Unidos. *Storm Surge Overview*. Online. Disponible: <https://www.nhc.noaa.gov/surge/>.
- [6] Tabla de mareas. *Tipos de mareas*. Online. Disponible : <https://tablademareas.com/mareas/tipos-mareas>.
- [7] S. Espín. *Efecto de las marejadas en las costas ecuatorianas*. Online. Disponible: <http://cidbimena.desastres.hn/docum/crid/Enero2006/CD-1/pdf/spa/doc9220/doc9220-contenido.pdf>, 2016.
- [8] F. Gonzalez. *Modelización estadística de eventos extremos de oleaje y nivel del mar*. 2013.
- [9] Meteorología y Estudios Ambientales Instituto de Hidrología. *Pronóstico de pleamares y bajamares en la costa caribe Colombiana*. Ministerio de Ambiente, Vivienda y Desarrollo Territoria Colombia., 2019.

- [10] F. Mancheno. *Las marejadas*. Online. Disponible: <https://lahora.com.ec/esmeraldas/noticia/1102211785/las-marejadas>, 2018.
- [11] Coria P. Martínez, A. *Distribución de probabilidad de la altura del oleaje dentro de la bahía de Todos Santos, B. C, México*. Ciencias Marinas, vol. 19, núm. 2, pp. 203-218, 1993.
- [12] Network Meteorology. *Douglas scale*. Online. Disponible: <https://www.meteorologiaenred.com/en/douglas-scale.html>.
- [13] Mauro Naghettini. *Fundamentals of Statistical Hydrology*. Springer., 2017.
- [14] Silva. R. *Análisis y descripción estadística del oleaje*. Instituto de ingeniería UNAM, 2005.
- [15] J. T. Shiau. *Return period of bivariate distributed extreme hydrological events*. Springer, 2003.
- [16] Roldan B. Torres I. *Distribución de Gumbel: Antecedentes y aplicación*. Miscelánea Matemática UNAM 69 pp. 111-125, 2019.
- [17] El Universo. *Cuatro cantones de Esmeraldas declarados en emergencia por oleajes*. Online. Disponible: <https://www.eluniverso.com/noticias/2018/12/27/nota/7114488/cuatro-cantones-esmeraldas-declarados-emergencia-oleajes/>, 2018.
- [18] Instituto Oceanográfico y Antártico de la Armada del Ecuador. *INOCAR*. Online. Disponible: <https://www.inocar.mil.ec/>.