

Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos. UNIVERSIDAD DE CANTABRIA

Evaluación de modelos de valores extremos sobre la cota de inundación y sus componentes

Trabajo realizado por:

Marco Hernández Brito

Dirigido:

Melisa Menéndez García Héctor Lobeto Alonso

Titulación:

Máster Universitario en Costas y Puertos

Santander, 13 de septiembre de 2023





Índice

1.	Intr	roducción	2
1	.1.	Objetivos del TFM	7
2.	Date	tos	7
2	.1.	Datos de nivel del mar	8
2	.2.	Datos de oleaje	8
3.	Zon	nas de estudio	8
3	.1.	Localizaciones	8
3	.2.	Características climáticas	9
4.	Met	todología	10
4	.1.	Preprocesado de los datos	10
4	.2.	Análisis de extremos: Obtención de periodos de retorno	13
4	.2.1.	Modelos basados en bloques de máximos	13
4	.2.1.1	1. Modelo Gumbel	13
4	.2.1.2	2. Modelo GEV	15
4	.2.1.3	3. Distribuciones mixtas	15
4	.2.2.	Modelos basados en excedencias sobre un umbral	17
	4.2.2	2.1. Modelo GPD – P	17
	4.2.2	2.2. Exponencial	
	4.2.2	2.2.1. Selección del umbral	19
	4.3.	Significancia de ξ	20
5.	Resi	sultados y Discusión	21
6.	Con	nclusiones	
Ref	erenc	cias	
Ane	exo		





Resumen

El nivel del mar junto con el oleaje son los principales indicadores que se emplean para entender el comportamiento de la superficie del océano. El entendimiento de su comportamiento es de gran importancia a la hora de prevenir los impactos que las condiciones extremas pueden generar en la costa. Ante este contexto, el presente trabajo pretende contribuir al entendimiento de las distintas variables que constituyen el nivel del mar total. Concretamente, se centra la atención en los distintos modelos estadísticos clásicos existentes que se utilizan para predecir el comportamiento de los valores extremos de estas variables, añadiendo un modelo más que, si bien no ha sido tan explorado, presenta visibles mejoras en los ajustes de estos valores. El estudio se aplica en 21 puntos costeros pertenecientes a distintas regiones climáticas. El propósito es analizar el desempeño de dichos modelos ante distintas condiciones extremas que presentan ciertas problemáticas a la hora de su modelado.

Abstract

Sea level and waves are the main indicators used to understand the behavior of the ocean surface. The understanding of their behavior is of great importance when it comes to prevent the impacts that extreme conditions can generate on the coast. In this context, the present work aims to contribute to the understanding of the different variables that constitute the total sea level. Specifically, attention is focused on the different existing classical statistical models that are used to predict the behavior of the extreme values of these variables, adding one more model that, although it has not been explored as much, presents visible improvements in the adjustments of these values. The study is applied to 21 coastal points belonging to different climatic regions. The purpose is to analyze the performance of these models under different extreme conditions that present certain problems when modeling them.





1. Introducción

Los fenómenos de inundación en la costa son generados por la acción de las ondas bajo determinadas condiciones de oleaje y nivel del mar. Sin embargo, esta afirmación es demasiado simplista como para comprender la verdadera complejidad del fenómeno. La aparición, intensidad y duración de los fenómenos extremos que inducen a inundaciones costeras provocadas por el oleaje son el resultado de una combinación de muchos procesos hidrodinámicos de distinta naturaleza, que operan a varias escalas espacio-temporales e interactúan entre sí. Además de esto, factores como la batimetría y las características estructurales y geométricas de las infraestructuras de protección pueden desempeñar un papel importante en la dimensión de estos sucesos. Por eso en esta primera sección introductoria se presenta un breve resumen de los distintos factores que contribuyen al nivel del mar y a la aparición de fenómenos extremos y sus impactos.

En primer lugar, el conocido en inglés como still water level (*SWL*), que se podría traducir al español como el nivel del mar "estático" o "en calma", indica la altura de la superficie del mar sobre un nivel de referencia, normalmente respecto a un datum de marea. Este se puede entender principalmente como la combinación de tres componentes: el nivel medio del mar, las mareas astronómicas y la marea meteorológica (*Caruso & Marani, 2022*).

(a) El nivel medio del mar (MSL) representa la altura del nivel del mar promediada sobre periodos de tiempo del orden de meses o años, alrededor de la cual la marea oscila. Este nivel se define como la altura media de la superficie del mar relativa a un punto de referencia fijo, realizándose el promediado abarcando un periodo de tiempo suficientemente largo para eliminar los efectos de la marea y los fenómenos atmosféricos sobre la altura de la superficie del mar. De no existir perturbaciones, el MSL debería coincidir con la superficie geopotencial conocida como geoide (*Weisse & von Storch, 2010*).

(b) La marea astronómica (AT) es la oscilación periódica de la superficie del agua alrededor del MSL, causada por la variación en el tiempo de la atracción gravitacional de la Luna y el Sol a medida que estos cambian su posición relativa respecto a la Tierra en rotación. Acorde a esto, la fuerza gravitacional resultante que actúa sobre los cuerpos de agua marinos en cualquier instante dado se debe a un elevado número de contribuciones, cada una con una frecuencia específica y asociada a un elemento astronómico específico perteneciente a la posición relativa de los elementos del sistema Tierra – Sol – Luna. Dichas contribuciones, comúnmente conocidas como componentes de la marea o constituyentes, abarcan un amplio rango de escalas temporales (frecuencias), incluyendo desde variaciones de menos de un día hasta de más de una década. De hecho, el número de componentes armónicas que se distinguen excede las 500 (*Godin, 1972; Pugh & Woodworth, 2014*).

A pesar de que el número de componentes armónicas es considerablemente elevado, sólo varias de estas generan una contribución significativa a la oscilación rítmica de la superficie del mar alrededor del nivel medio. Por esto, no es notablemente difícil identificar el origen de ciertos ciclos mareales de interés práctico para entender los patrones de variación del rango mareal, el cual se define como la diferencia vertical en altura entre las mareas altas y bajas consecutivas. Estos corresponden a los puntos más alto y bajo alcanzados durante un día lunar (24 h 50 min; el periodo de rotación aparente de la Luna alrededor de la Tierra), que pueden fluctuar desde centímetros hasta más de 15 metros dependiendo de la localización.

Debido al elevado número de constituyentes, la elevación de la superficie resultante es una complicada, aunque predecible, función de tiempo que puede ser expresada como la superposición de un largo número de constituyentes armónicos:





$$\xi(t) = A_0 \sum_{i=1}^{N} A_i \cos(\omega_i t - \varphi_i)$$
⁽¹⁾

donde A_0 es el nivel medio del agua sobre el nivel de referencia (datum); N es el número total de constituyentes; i refiere a un constituyente particular; A_i , ω_i y ϕ_i representan su amplitud, frecuencia angular y fase relativa respectivamente. En otras palabras, para predecir el nivel de marea es necesario conocer la frecuencia, amplitud y fase de cada componente considerada en la expresión (1). No obstante, las ondas de marea tienen una longitud de onda muy larga, mucho mayor que la profundidad del océano, por lo que se propagan en condiciones de aguas poco profundas. Por lo tanto, cuando alcanzan profundidades costeras reducidas son frenadas, amplificadas, moduladas y distorsionadas por diferentes mecanismos hidrodinámicos no lineales, entre los que se incluyen las fuerzas de fricción y la interacción con el fondo y los contornos de la costa. Los diversos efectos de la propagación de la onda de marea en aguas poco profundas incluyen la generación de armónicos de orden superior y, en particular, la modificación de la amplitud y la fase de las componentes de las mareas. Es decir, que la amplitud y el momento de la marea no sólo están controlados por factores astronómicos, sino que la hidrodinámica local también desempeña un papel importante.

El cálculo de amplitudes y fases en un punto dado del océano, usando registros de nivel del mar extensos en el tiempo y considerando un número apropiado de constituyentes, permite realizar predicciones precisas de la marea para dicha localización. Con estos términos, conocer la amplitud de los 4 principales constituyentes de la marera (lunar – solar (K1), lunar diurna principal (O1), luna semidiurna principal (M2) y solar semidiurna principal (S2)) hace posible determinar el carácter de la marea en un área determinada mediante la siguiente relación entre las amplitudes de los armónicos diurnos y semidiurnos, conocida como el número de forma de la marea (F, eq. 3) (*Pugh & Woodworth, 2014*),

$$F = \frac{A_{K_1} + A_{O_1}}{A_{M_2} + A_{S_2}}$$
(2)

Dependiendo del valor de F, el régimen mareal se puede clasificar como semi-diurno (0 < F < 0.25), mixta predominantemente semidiurna (0.25 < F < 1.50), mixta predominantemente diurna (1.50 < F < 3.00) y diurna (F < 3.00).

(c) El *residuo meteorológico (NTR)* es la perturbación en la elevación nivel del mar causada por las condiciones meteorológicas en escalas de tiempo que varían de minutos a días. En particular, la presencia de sistemas atmosféricos de bajas presiones en áreas costeras puede resultar en fuertes y persistentes vientos soplando hacia tierra. El estrés ejercido por estos vientos sobre la superficie del mar puede inducir sobre el este un set-up por viento en la zona costera. Este efecto es especialmente importante cuando el viento sopla hacia la costa, debido a la disminución progresiva de la profundidad y a la presencia de la línea de costa que actúa como un contorno sólido, pudiendo forzar un aumento del nivel de mar a niveles significativamente altos. Por otra parte, la disminución local de la presión atmosférica asociada a una tormenta puede conllevar en sí una subida del nivel del mar como respuesta. Este fenómeno se conoce generalmente como efecto del barómetro inverso, ya que un descenso de 1 mbar en la presión atmosférica provoca un aumento de 1 cm en el nivel del mar. No obstante, por lo general, la contribución al cambio del nivel del mar debida a los descensos de la presión atmosférica es significativamente menor que la causada por la acumulación de agua contra la costa por la acción del viento. En conjunto, el incremento del nivel del mar asociado a tormentas depende de factores locales tales como la configuración de la costa y la profundidad, más allá de la severidad de las condiciones atmosféricas. Como resultado, el incremento del





nivel del mar promedio puede encontrarse dentro de un rango que abarca desde los centímetros al metro, aunque, bajo condiciones excepcionales, se pueden alcanzar elevaciones de varios metros (*Pugh, 1987*). Es interesante destacar que estos cambios inducidos meteorológicamente en el nivel del agua a lo largo de la costa no se producen necesariamente bajo circunstancias de tormenta, sino que también pueden ocurrir con condiciones de viento más moderadas. Sin embargo, el término genérico utilizado para denominar este fenómeno es el que se emplea en inglés como *storm surge* (*SS*).

La naturaleza de este fenómeno, teniendo en cuenta los factores que lo generan, hace que esta contribución a la variación del nivel del mar posee un carácter aleatorio, al contrario que la contribución periódica de la AT. En consecuencia, dado que según la ecuación (2) el nivel del mar en cualquier momento resulta de la superposición de las contribuciones astronómica y meteorológica, esta última puede considerarse como una variabilidad aleatoria superpuesta a la AT. Por lo tanto, asumiendo esta hipótesis como correcta, la señal del SS (o NTR), se puede obtener simplemente restando el nivel de marea previsto del nivel del mar medido. Respecto a esto, cabe señalar que la definición de la componente meteorológica supone que la AT y el SS son independientes entre sí. Sin embargo, en muchos casos, una simple extracción de la predicción de AT a las alturas del nivel del mar registradas no es correcta, ya que ambos fenómenos interactúan entre sí de forma no lineal (*Rodríguez et al, 1999*).

Con lo comentado hasta ahora, las tres anteriores componentes (MSL, AT y NTR) definen el SWL, considerando como la elevación de la superficie del mar asumiendo la ausencia de las olas. En otras palabras, la superposición de estos 3 factores determina el nivel de referencia sobre el que el oleaje generado por el viento se propaga. En línea con esto, las fluctuaciones en la superficie libre debido a la secuencia de crestas y valles inherentes al oleaje cuando este se propaga por zonas de aguas someras pueden conducir a variaciones en el SWL. Estas se deben tener en cuenta a la hora de identificar el nivel medio de la superficie del agua resultante, que es conocido como nivel medio del agua o, en inglés, *mean water level (MWL)*, sobre el cuál aparecen las fluctuaciones de pequeño periodo intrínsecas al oleaje de viento. El oleaje por viento es el responsable último en el riesgo asociado a los eventos extremos de inundación. Por este motivo también parte de este documento se destina a describir las características básicas del oleaje, así como las variaciones del nivel del mar que este puede inducir.

Como se acaba de comentar, las olas generadas por el viento, comúnmente conocidas como *oleaje de viento*, son fluctuaciones de las superficie del mar sobre el SWL, las cuáles son generadas por la transferencia de energía y momento que realiza el viento turbulento a la interfase aire-agua. En este sentido, es interesante denotar que, al igual que con el SS, el oleaje de viento representa la respuesta del mar a la acción del viento y las variaciones de presión en la interfase atmósfera-océano, si bien a escalas espacio-temporales significativamente diferentes y a partir de mecanismos físicos distintos. En este sentido, ambos fenómenos tienen una naturaleza aleatoria, especialmente en el caso del oleaje de viento, cuya génesis está estrechamente relacionada con la turbulencia en el límite marino-atmosférico.

Durante su propagación hacia la zona de rotura, de aguas profundas a someras, el oleaje contiene energía y momento. Durante los procesos de rotura, la energía del oleaje se disipa, pero el momento se transfiere a la columna de agua. Las variaciones horizontales en el flujo de momento hacia la costa actúan como una fuerza sobre la masa de agua que se traduce en un incremento en el MWL conocido como *setup* (η); una variación media del MWL por la rotura la del oleaje (*Dean & Walton, 2009; Svendsen, 2006*).

A la luz de lo descrito hasta ahora, la aparición, intensidad y duración de los fenómenos extremos que inducen a inundaciones costeras surgen de combinaciones concretas de AT, SS, así como del η . Sin embargo, es importante mencionar que la acción conjunta de estos factores no es una simple superposición lineal, sino que interaccionan entre ellos de manera que el potencial efecto individual de cada uno puede





generar variaciones significativas (*Sundar, 2021; Parker, 2007; Idier, et al., 2019*). En esto reside la complejidad del proceso, en su carácter aleatorio multivariado y no lineal, constituyendo un desafío considerablemente difícil determinar el grado de dependencia entre cada par de contribuciones individuales (*Rodríguez et al., 1999*)

A la hora de analizar los fenómenos extremos en estadística, se consideran como valores extremos aquellos situados en la cola de las distribuciones de probabilidad; cuya probabilidad de ocurrencia es considerablemente baja y decreciente a medida que aumenta su magnitud. La predicción de eventos extremos es un desafío a día de hoy, ya que requiere la existencia de sucesos previos en los registros. Además de esto, interesante observar cómo se ven afectados por factores externos, tal y como es el caso de las inundaciones costeras y su vinculación con el cambio climático, sobre todo para los fenómenos que involucran mayores escalas de tiempo (Coles, 2001; Vega, 2018).

En el contexto actual, las edificaciones y construcciones de carreteras a lo largo de la costa son una opción muy frecuente, y casi obligatoria, para el desarrollo de las poblaciones y el transporte de mercancías y personas en muchas zonas. Por otro lado, teniendo en cuenta la situación privilegiada de los terrenos ocupados por las carreteras costeras, es muy común la existencia de espacios públicos abiertos considerados de alto valor social por muchos tipos de usuarios situados entre la carretera y el mar, sobre todo en ciudades densamente pobladas (*Macdonald, 2018*).

Por esto y considerando todo lo anterior, la proximidad al mar de dichos asentamientos los convierte en un escenario de riesgo en el que infraestructuras, peatones y vehículos son susceptibles de experimentar los impactos negativos de los peligros asociados a los procesos hidrodinámicos que se generan bajo determinadas condiciones extremas (*Smith & Juria, 2019*). En este sentido, es importante resaltar la zona cercana a la costa como un área altamente energética donde tierra, mar y atmósfera convergen e interactúan a diferentes escalas espacio-temporales, induciendo una dinámica extraordinariamente compleja, que puede causar daños potenciales en edificaciones y carreteras, interrupciones de la cadena de suministro (con las pérdidas socioeconómicas asociadas) e incluso víctimas (*Chen et al., 2007; Knott et al 2017*). Es por esto que el estudio de valores extremos es de gran importancia en muchas áreas de trabajo, como pueden ser las geociencias y la ingeniería, permitiendo estimar los tiempos de retorno de los eventos extremso y de esta manera medir y tratar de minimizar los riesgos asociados a escenarios desfavorables.

Uno de los principales objetivos en los análisis de frecuencias de ocurrencia es el de calcular los intervalos promedio de recurrencia del periodo de retorno. Este es un concepto frecuentemente usado en los análisis de riesgo hidrológico y geofísico. En los procesos estacionarios, el periodo de retorno (T_R) de la magnitud de un evento se define como el tiempo medio que transcurre entre dos excedencias consecutivas de dicha magnitud. En otras palabras, se podría decir que la magnitud, de media, se espera que sea excedida en cada periodo de retorno. Si el máximo anual X es excedido de media una vez en T_R años, su probabilidad de excedencia [E(X) = 1 - G(X)] en un año determinado será:

$$E(X) = P_{H \ge X} = \frac{1}{T_R(X)}$$
(3)

Si representamos cada periodo de retorno respecto a la magnitud del evento correspondiente (*nivel de retorno*), obtenemos lo que se conoce como gráfico de periodos de retorno (*Figura 1*). El desafío reside en estimar qué valores (p. Ej. SL) se podrían alcanzar en los próximos 100 años a partir de registros de no más de 30 años. Es aquí cuando adquiere valor la Teoría de Valores Extremos (*TVE*), descrita en el apartado 4.2, ya que esta nos permite realizar estimaciones a periodos de retorno largos a pesar de no disponer de sucesos extremos con magnitudes equivalentes en las series. Estos gráficos son particularmente convenientes tanto para representar los ajustes de modelos como para su validación, debido a su facilidad





de interpretación y que permiten representar la cola de la distribución de manera que el efecto de la extrapolación se resalta (*Coles, 2001*). La complejidad reside en no disponer de series de datos lo suficientemente largas como para poder hacer extrapolaciones robustas, ya que se necesita un mínimo número de registros en las muestras de valores extremos para que estas se puedan considerar representativas.

Una de las principales problemáticas que se pretende abordar con este trabajo es el error que genera en los ajustes de las muestras los valores extremos que tienen distinta génesis. Es decir, en las zonas que se encuentran bajo la influencia de fenómenos atmosféricos, como los ciclones tropicales, se generan valores etremos incluso más inusuales que los ya de por sí existentes. Esto pone de manifiesto en estas gráficas la



Figura 1. Periodos de retorno asociados a una muestra de máximos anuales con valores ajustados con modelos GEV (negro) y Gumbel (gris) en la zona de Galveston, Texas. Resaltadas en azul y rojo las dos familias de extremos. Se puede ver cómo el ajuste del modelo GEV empeora a medida que aumenta el periodo de retorno de la estimación.

existencia de distintas familias de extremos que seguirían ajustes diferentes, ya que no tienen un origen común (*Figura 1*).

A esto habría que añadirle una segunda problemática que, si bien aquí no se pretende abordar, contribuye a la complejidad del problema. Debido a que vivimos en una época en la que el clima está siendo sustancialmente modificado por la actividad humana, los cambios en el MSL pueden tener un rol importante modificando la frecuencia y magnitud de los niveles extremos del mar (*Figura 2*) (*IPCC, 2007a*), aumentándola junto con su probabilidad de ocurrencia. Esto hace que las condiciones cambiantes a futuro del MSL debido a su aumento como efecto directo del cambio climático (*MSLR*) sean un factor más a tener en cuenta en los estudios sobre el clima.

A partir de las diferentes proyecciones de MSLR, para distintos escenarios de emisiones, siguiendo resultados del 4^{to} Informe de Evaluación del IPCC (*IPCC*, 2007b), se han obtenido estimaciones de 0.79 m de MSLR global para el año 2095 (respecto a 1990). Si esto se representa en un gráfico de periodos de retorno se ve que, efectivamente, aumenta la frecuencia de ocurrencia de valores que anteriormente, de media, sólo se registraban cada 50 años, para 2095 sería cada menos de 25 años (*Hunter*, 2010).







Figura 2. Periodos de retorno asociados a una muestra de máximos anuales con valores ajustados con modelos GEV y Gumbel para la proyección de 0.79 m de MSLR del IPCC.

Con las consideraciones anteriores en mente, la motivación de este estudio es la comparación y mejora del entendimiento del uso de distintos modelos estadísticos de extremos, de manera que se permita discernir las características bajo las cuales unos sean más adecuados que otros, incluyendo los casos en los que el comportamiento de las distintas familias de valores extremos que en una misma localización pueden tener lugar. Para ello, en este estudio se ha escogido una serie de localizaciones distribuidas a lo largo de distintas regiones del planeta; regiones que se encuentran o bien afectadas por la acción de huracanes / tifones / ciclones o bien por ciclones extra-tropicales, con el fin de evaluar el desempeño de las distintas metodologías.

Este estudio se ha estructurado en 7 secciones, incluyendo esta introducción. Los datos usados para llevar a cabo la investigación y las zonas de estudio están descritos en las secciones 2 y 3, respectivamente. La metodología seguida para alcanzar los objetivos propuestos se indica en el apartado 4. Los resultados obtenidos en el estudio se presentan y discuten en las secciones 5 y 6. Finalmente, las conclusiones alcanzadas en este trabajo están resumidas en la sección 7.

1.1. Objetivos del TFM

El objetivo de este trabajo académico está enfocado en el entendimiento del comportamiento de los valores extremos de nivel del mar y de oleaje mediante diferentes técnicas de modelado de extremos, con especial énfasis en las zonas afectadas por ciclones tropicales. Se pretende evaluar qué modelos estadísticos de extremos son más idóneos para cada zona/variable analizada, de manera que esto facilite la elección del más conveniente a la hora de caracterizar las condiciones extremas de las variables que en este estudio se incluyen.

2. Datos

El estudio se ha llevado a cabo usando dos tipos distintos de datos. Por un lado, se han empleado datos observados de nivel del mar, obtenidos a partir de medidas directas en las distintas zonas de estudio. Por el





otro, se hizo uso de información simulada procedente de modelos numéricos para la variable de oleaje. Las características más destacadas de estos conjuntos de datos se describen en los siguientes subapartados.

2.1. Datos de nivel del mar

Se emplearon registros de SL de diversos mareógrafos localizados en los distintos puntos de estudio, los cuáles era necesario que contuviesen largas series temporales de datos horarios continuos. A dicha red de mareógrafos se tuvo acceso a través del portal web del proyecto Global Extreme Sea Level Analysis (GESLA, concretamente de su tercera versión GESLA-3) la cual forma parte del Global Sea Level Observing System (GLOSS). Si bien, este conjunto de datos cuenta con más de 5.000 registros, cubriendo las zonas costeras de alrededor de 90 países a lo largo de todo el planeta, la distribución de estos mareógrafos es particularmente escasa en algunas regiones (p. ej., este de África) o, en otras, incluso no tienen presencia (p. ej. India). También cabe destacar que el hemisferio sur presenta una cobertura geográfica más pobre que la del hemisferio norte y con registros más cortos. Con todo esto, dicha información resulta esencial para la investigación de cambios del nivel del mar y circulación oceánica. Dada su amplia distribución y alta frecuencia en los datos, estos registros son más que adecuados para proporcionar la información del nivel del mar que en este trabajo es requerida (Haigh et al., 2021; Woodworth et al., 2017). Las series seleccionadas contenían datos de al menos 1979 en adelante, albergando todas como mínimo 40 años de datos que, atendiendo a las recomendaciones de la World Meteorological Organization, la cual sostiene que son necesarios un mínimo de 30 años de registros para los estudios climáticos, estas pueden ser vistas como suficientemente extensas para obtener información de su variación a lo largo de las últimas décadas (Semedo et al., 2010).

2.2. Datos de oleaje

La información necesaria para analizar la sobreelevación del nivel del mar en la costa fue obtenida a partir de registros de oleaje. Dicha información se ha recogido de datos generados sintéticamente a partir del modelo numérico GOW2, cuyas salidas se asociaron a los puntos de la malla más cercanos a la ubicación de los mareógrafos. Este modelo es un Hindcast de oleaje global, creado a partir del modelo WAVEWATCH III, que proporciona datos de oleaje globales y homogéneos en las zonas continentales costeras e islas oceánicas de 1979 en adelante. Al incorporar varias mejoras respecto a las versiones anteriores, en los forzamientos, parametrizaciones y configuración del grid, genera información fiable para realizar análisis de valores extremos, como es el caso de este estudio (*Menéndez & Losada, 2017*).

3. Zonas de estudio

3.1. Localizaciones

Se realizó una selección de diversos puntos localizados en distintas zonas a lo largo de la costa (*Figura* 3). Como en este estudio se pretende comparar la capacidad de modelar las distintas familias de extremos, dichos puntos fueron escogidos atendiendo a los patrones climáticos presentes en sus regiones. A su vez, se mantuvieron ciertos criterios de calidad de los datos que los dispositivos de medida ofrecían (una mínima continuidad de los datos, mínima cobertura temporal y espacial, un mínimo de 60% de datos efectivos), considerando las limitaciones descritas en el apartado 2.1. Una vez terminada la selección de puntos que cumplían con los criterios de calidad, y después de descartar algunos durante el desarrollo del análisis, se seleccionaron un total de 21 mareógrafos.

Los puntos 1 al 4 se encuentran en la zona de Golfo de México (1 a 3) y el Caribe (4). Del punto 5 al 8 se cubre la costa este de Norteamérica, mientras que los puntos 20 y 21 la costa oeste de Norteamérica. Con el punto 9 se recoge información en el continente europeo, en concreto en la zona del Mar del Norte. Los





puntos 10 a 14 representan la zona del este sudeste asiático. Los puntos del 15 al 18 cubren la zona de Oceanía incluyendo la Polinesia (*Figura 3*).



Figura 3. Localizaciones de los puntos de estudio.

3.2. Características climáticas

Respecto a las características climáticas predominantes en las distintas regiones, se puede hacer una distinción principal: zonas bajo influencia de ciclones tropicales y zonas con influencia de ciclones extratropicales. En las zonas de Europa, sur de Oceanía, este de Sudamérica y noroeste de Norteamérica, los valores extremos están claramente dominados por las tormentas extra-tropicales; mientras que, en el resto de regiones, los eventos más extremos de los registros están generados por ciclones tropicales (*Muis, et al. 2016*). En este trabajo, los puntos 9, 13, 17, 20 y 21 corresponden a las zonas con influencia de ciclones extra-tropicales, mientras que el resto de puntos, se encuentran en zonas que se pueden ver afectadas por ciclones tropicales (*figura 3*).

A modo descriptivo, en la *figura 4* se ha representado el rango de la marea astronómica (RM) para cada ubicación. Se puede observar una variación notable de unas zonas a otras, que comprende desde las que presentan un régimen micromareal (RM < 2 m), como es el caso de la zona del Golfo de México (más resguardada de la onda de marea); hasta las que presentan un régimen macromareal (RM > 4 m), como sucede en el Mar del Norte (4.13 m); pasando por las mesomareales (2 m < RM < 4 m) en el oeste de Norteamérica (2.18 m) o en costa sureste de Japón (2.13 – 2.10 m).







Figura 4. Rango de marea en los 21 puntos de estudio.

Por otro lado, con las series históricas del SL podemos ver la evolución del nivel del mar a lo largo del tiempo. En la *figura 5* se han representado dos ejemplos de zonas en las que hay presencia de ciclones tropicales y zonas bajo influencia de ciclones extra-tropicales. Rápidamente se puede una clara diferencia tanto en la frecuencia como en la intensidad de los eventos extremos de SL. En las zonas extra-tropicales (puntos 9 y 17) los picos extremos de la serie son recurrentes y, por lo general, no se observan valores que destaquen demasiado que otros por encima de los 8 y 2 m en los paneles c y d respectivamente. Por el contrario, en los paneles a y b (correspondientes a los puntos 3 y 18 respectivamente) sí se aprecian picos que sobresalen por encima de los demás eventos, doblando en algunos casos las alturas de otros extremos. La frecuencia con la que estos aparecen también es notablemente menor que la de los picos en zonas extra-tropicales.

Adicionalmente, se ha añadido la evolución del valor de la mediana anual calculada a partir de los valores horarios de las series con el fin de resaltar la tendencia a largo plazo en el MSL. Se muestra en rojo la recta de ajuste por mínimos cuadrados de la mediana anual a un polinomio de grado uno.

4. Metodología

4.1. Preprocesado de los datos

Una vez obtenidas las series descritas en el apartado 2, se lleva a cabo una revisión de la calidad de los datos para eliminar anomalías en las series. Además, la propia fuente de los datos realiza un control de calidad previo a su distribución, ajustando también las fechas de cada registro a *Coordinated Universal Time (UTC)*. Con todo ello, sólo un número reducido de valores anómalos fueron descartados para los siguientes pasos del análisis.

De antemano, es relevante aclarar que tres son las variables meteoceánicas que se analizan en este trabajo: el NTR, el setup (η) y el TWL. El análisis del TWL aporta información de relevancia respecto al riesgo de inundación, mientras que, tanto el NTR como el η proporcionan información sobre los forzamientos







Figura 5. (Galveston, Golfo de México; Papeete, Polinesia) y X (Cuxhaven, Europa; Wellington, Oceanía)





climáticos de cada zona.

4.1.1. Análisis armónico

Para obtener las series temporales de las variables de interés, en primer lugar, se realizó una descomposición armónica de la serie de SL con la que se extrajeron las amplitudes y frecuencias de las principales componentes de la marea. El análisis armónico se ha realizado haciendo uso de la herramienta de análisis mareal U-tide (*Codiga, 2011*), cuya base consiste en asumir que las variaciones de la marea pueden ser representadas a partir de un número finito, N, de términos armónicos de la forma que se presenta en la expresión (1). Cabe mencionar que, previo al análisis, se eliminó la tendencia del registro de SL restando el valor de cada mediana anual a la serie.

Como ya se ha comentado en la descripción de la AT, la amplitud y el momento de la marea no sólo están controlados por factores astronómicos, sino que la hidrodinámica local también desempeña un papel importante. En consecuencia, mientras que las frecuencias de las componentes armónicas son bien conocidas, las amplitudes y las fases de dichos constituyentes varían con la localización y por tanto se deben determinar a partir de la medida instrumental de SL de cada una de las ubicaciones analizadas. El objetivo del análisis armónico de la marea es minimizar los residuos (ϵ) en la expresión (4) para estimar estos parámetros (*Godin, 1972*).

$$\xi(t) - \left[A_0 \sum_{i=1}^{N} A_i \cos(\omega_i t - \varphi_i)\right] = \varepsilon_i$$
(3)

incluyendo dichos términos residuales (NTR) posibles contribuciones de origen "no astronómico", inducidos por forzamientos meteorológicos. De esta manera, al ser una variable totalmente determinista, se puede reconstruir la serie de AT, mientras que el NTR (ϵ) se calcula substrayendo la AT al SL.

No obstante, es importante mencionar que la secuencia de SL está estrechamente relacionada con procesos no estacionarios, debidos a tendencias a largo plazo en el MSL, la AT determinista y las interacciones entre la AT y el NTR. Estas interacciones entre la marea y el storm surge pueden modificar la amplitud y las fases del NTR, sobre todo en zonas costeras (*Caruso & Marani, 2022*). Por lo tanto, a pesar de que no se haya considerado en este trabajo, dicho efecto debe ser tenido en cuenta cuando se separan las componentes de la AT y el NTR. Esto bien puede llevar a imprecisiones en la estimación del NTR, ya que, si bien el SL se asume como la suma de la AT con el NTR, estas dos componentes son independientes e interactúan entre sí de forma no lineal, sobre todo en zonas cercanas a la costa, tal y como se comentaba en la parte introductoria del documento.

4.1.2. Estimación del η y cálculo del TWL

En segundo lugar, para el cálculo del η se emplearon los datos de oleaje descritos en el apartado 2.2. Para esto se siguió la formulación propuesta por *Stockdon et al.*, (2006), con la que se realiza una estimación empírica de esta variable de forma paramétrica, a partir de la siguiente expresión.

$$\eta = 0.35\beta_f \cdot (H_0 L_0)^{1/2} \tag{4}$$

donde β_f es la pendiente media del fondo; H₀ la altura de ola en aguas profundas (obtenida a partir de las series temporales del modelo GOW) y L₀ es la longitud de onda en aguas profundas calculada a partir del periodo medio.





Por último, para realizar el cálculo de las series de TWL, como este es el resultado final de la combinación de AT, NTR y η , simplemente se realizó la suma del η estimado a los registros originales de SL.

4.2. Análisis de extremos: Obtención de periodos de retorno

A partir de la TVE se analiza y desarrolla el estudio de las probabilidades relacionadas con las colas de las funciones de distribución. Esta teoría engloba diferentes modelos de extremos y cada uno se basa en la función de distribución de extremos que utiliza.

Existen dos aproximaciones a los modelos de extremos: modelos exactos y modelos asintóticos. Debido a que los primeros no presentan gran aplicabilidad en los casos de estudio reales, estos trabajos se centran en los dos modelos básicos que permiten resolver el modelo asintótico de extremos. Dichos modelos derivan de diferentes teoremas de la TVE:

- I. El Teorema de Fisher-Tippet, en el que se basa la aproximación clásica de valores extremos (Galambos, 1987).
- II. El Teorema de Pickands, a partir del que se define la aproximación basada en el ajuste de la cola de la distribución (*Pickands, 1975*).

Estas dos aproximaciones se diferencian eminentemente por el método de selección de los datos de partida; mientras que la primera considera todo el rango de valores de la muestra de la población, la segunda sólo considera las sección más alta de valores, aquellos que superan un cierto valor umbral.

Los modelos de extremos se construyen a partir de unas hipótesis de partida y un desarrollo matemático, de forma que estos permiten realizar extrapolaciones a valores que no se encuentran dentro de los datos observados junto con sus probabilidades aproximadas de ocurrencia.

Los modelos de extremos básicos que se derivan de cada teorema se describen a continuación:

4.2.1. Modelos basados en bloques de máximos

El teorema de Fisher-Tippett demuestra que si la función de distribución común (F_x), de una secuencia de observaciones aleatorias e independientes ($X_1, X_2, ..., X_n$), se encuentra dentro del máximo dominio de atracción [$F \in MDA(G)$] de una función de distribución no degenerada (G), entonces G pertenece a alguna de las familias de distribuciones que engloba la Función Generalizada de Extremos (GEV, *Generalized Extreme Value*), descritas con mayor detalle en el siguiente apartado.

El desarrollo de este modelo está pensado para trabajar con datos registrados en intervalos de observaciones de longitud n, resultando esto en que los valores de la muestra de extremos que se utilizan para ajustar a la función GEV sean los asociados a cada valor máximo de cada uno de dichos bloques. Dichos bloques están asociados normalmente a periodos de tiempo que, para variables geofísicas, suelen ser anuales, para los cuáles n será el número de observaciones realizadas en un año y el máximo de un bloque corresponde a un máximo anual.

4.2.1.1. Modelo Gumbel

Tanto la expresión de la función de distribución acumulada como la de densidad de probabilidad están compuestas por una serie de parámetros que son los que determinan el tipo de familia a la que pertenecen:

- > Parámetro de localización: μ
- > Parámetro de escala: ψ





> Parámetro de forma: ζ

El primero representa (μ) los valores medios de la variable que se estudia e identifica el valor asociado a la probabilidad de no excedencia exp(-1). Con el segundo parámetro (ψ) se controla la dispersión de las medidas de la variable dentro del intervalo de la distribución, permitiendo esto contraer o expandir el rango de esta a lo largo del eje de abscisas. El último parámetro (ζ) es el que condiciona el tipo diferente de distribución. Para $\zeta \neq 0$ se engloba dos familias de distribución de valores extremos (tipos I y II): Fréchet para el caso de $\zeta > 0$ y Weibull para el caso de $\zeta < 0$; mientras que el caso para $\zeta = 0$ se obtiene a partir del límite de la ecuación definida para Fréchet/Weibull cuando $\zeta \rightarrow 0$, definiendo así la familia de distribuciones tipo Gumbel cuya función de distribución (*CDF*) se expresa mediante:

$$G(x;\mu,\psi) = \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\psi}\right)\right]\right\}; \quad \text{para } \xi = 0 \tag{4},$$

mientras que su función de densidad de probabilidad (PDF) mediante la expresión:

$$g(x;\mu,\psi) = \frac{1}{\psi} exp\left(-\frac{x-\mu}{\psi}\right) \cdot exp\left[-exp\left(-\frac{x-\mu}{\psi}\right)\right]; \quad \text{para } \xi = 0 \tag{5},$$

siendo x los valores seleccionados para la muestra de extremos.

Estos tres tipos de familias que se ha mencionado que siguen el teorema de Fisher-Tippett tienen distinto comportamiento en la cola de la función de distribución (*figura 6.b*). Por lo general se emplean los términos de cola acotada para la familia de tipo Weibull, cola ligera para la de Gumbel y cola pesada para el caso de Fréchet. Esta distinción se explica a partir del aspecto que G adquiere en su valor más grande, x⁺. Si x⁺ < ∞ la familia será tipo Weibull que, para aplicaciones con variables geofísicas, por ejemplo, suele darse cuando las condiciones de contorno explican un lógico valor máximo a partir de un límite físico. Si x⁺ = ∞ , la familia de distribuciones será, o bien tipo Fréchet, o tipo Gumbel. Para el caso de Gumbel la cola de la distribución decae exponencialmente, dando lugar a las colas ligeras anteriormente mencionadas. El tipo Fréchet se da cuando la cola de la distribución decae como una función de potencia. Diferentes estudios han puesto en evidencia que variables en hidrología suelen comportarse con distribuciones de extremos de colas pesadas (*Katz et al., 2002*).



Figura 6. PDF (a) y CDF (b) de la GEV para los distintos parámetros de forma.





La aproximación clásica en la aplicación del modelo de extremos siempre ha consistido generalmente en considerar uno de los tres tipos de familias, siendo las expresiones particulares para cada una las siguientes:

- ► Tipo Gumbel: $G(z) = \exp[-exp(-z)]; -\infty < z < \infty$ (6)
- > Tipo Fréchet: $G(z;\xi) = \begin{cases} 0 & \text{si } z < 0\\ exp(-z^{-1/\xi}) & \text{si } 0 < z < \infty \end{cases}$ (7)
- > Tipo Weibull: $G(z;\xi) = \begin{cases} \exp\left[-(-z)^{1/\xi}\right] & \text{si } -\infty < z < 0 \\ 0 & \text{si } z > 0 \end{cases}$ (8)

donde $z = \frac{x-\mu}{\psi}$ es una variable aleatoria adimensional y ξ es el parámetro de forma ya descrito.

4.2.1.2. Modelo GEV

En estas aplicaciones de las tres familias de distribución el procedimiento consistía en adoptar una y estimar posteriormente sus parámetros de distribución. Sin embargo, esta metodología presentaba una serie de inconvenientes, ya que por un lado dicho criterio de selección se basaba únicamente en razones descriptivo-cualitativas y, por el otro, este no permitía estimar la incertidumbre resultante asociada a dicha selección en los cálculos posteriores. En esto reside la utilidad de la generalización de los tres tipos de familias que se muestra en la expresión (X.x). La unificación de las tres familias de distribución de valores extremos originales en una sola familia simplifica considerablemente su implementación estadística, permitiendo que el comportamiento de los valores extremos pueda adaptarse a un mayor rango en los casos en los que se pretenda realizar aplicaciones reales. A partir de la inferencia de ζ se consigue el comportamiento más apropiado de la cola, sin necesidad de realizar un juicio subjetivo previo para determinar qué familia individual de valores extremos adoptar. Además de esto, la incertidumbre en el valor inferido de ζ representa directamente la falta de certidumbre sobre cuál de las tres familias originales es más apropiada para un conjunto de datos dado. Esto resulta en que la función de distribución de la GEV se describe mediante la expresión:

$$G(x;\mu,\sigma,\xi) = exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\psi}\right)\right]^{-1/\xi}\right\} \quad \text{para } \xi \neq 0 \tag{9}$$

siendo la expresión para $\xi = 0$ la correspondiente a la función de distribución para la familia de distribuciones tipo Gumbel señalada en el apartado anterior (X.x), mientras que la función de densidad:

$$g(x;\mu,\sigma,\xi) = \frac{1}{\psi} \left[1 + \xi \left(\frac{x-\mu}{\psi} \right) \right]^{-(1+1/\xi)} \exp\left\{ - \left[1 + \xi \left(\frac{x-\mu}{\psi} \right) \right]^{-1/\xi} \right\} \quad \text{para } \xi \neq 0 \tag{10}$$

4.2.1.3. Distribuciones mixtas

En línea con los modelos de bloques de tiempo descrito, tanto la función de distribución como la función de densidad se pueden expresar como una combinación de otras funciones de distribución y densidad respectivamente (eq. 11). Esto es lo que se conoce como distribuciones mixtas, cuyas distribuciones individuales que las conforman reciben el nombre de *componentes de la mezcla (mixture components)*. Dicha combinación se realiza aplicando una suma con distintos pesos no negativos asignados, los cuáles por obligación deben sumar el valor de una unidad, ya que son las probabilidades asignadas a cada una de las componentes, los *pesos de la mezcla (mixture weights)*.





 $PDF_{mixta} = p_1 \cdot PDF_1 + p_2 \cdot PDF_2 + \dots + p_i \cdot PDF_i$ (11)

donde p_i es el peso asociado a cada una de las distribuciones.



Figura 6. PDF resultante (línea negra) de TWL y sus components de mezcla. La Segunda GEV no se puede observar con claridad debido a la escala del eje x. La línea roja indica la PDF empírica (Kernel). Primer gráfico: *Galveston Pier 21*. Segundo gráfico: *La Jolla* (Fuente: *SERDP*).

Para este estudio se seleccionaron como componentes de mezcla una distribución normal y dos GEV. La distribución normal se aplica a lo largo de toda la función de densidad para modelar las zonas medias y bajas de la misma. La primera GEV sí se aplica limitando su µ por encima del valor de la mediana de la muestra de extremos (percentil 50), para modelar así de forma más precisa la parte superior de la distribución. Mientras que la segunda GEV se restringió por encima del percentil 99.95, con el fin ya sí de modelar los valores de los eventos extremos más inusuales registrados.

Por otra parte, es relevante considerar que, a la hora de aplicar el ajuste a las muestras de las distintas variables, al no existir un forzamiento común a todas y sobre todo proceder de distintas ubicaciones, los valores más adecuados para los parámetros de las distribuciones no serán los mismos para cada variable en cada punto. Es por esto por lo que se ha optado por realizar los ajustes estableciendo unos límites de rango en los parámetros de localización (Tabla 1). De esta manera, se consideraba como el ajuste óptimo aquel que maximizaba la verosimilitud, estimada a partir de una función de mínimos aplicada con un solver de programación no lineal, obteniendo una mejora visible como se muestra en el ejemplo de la sección 5.2 (*figura 11*).

Tabla 1.	Valores a	los que se	restringe el	parámetro μ	para cada GE	V.
----------	-----------	------------	--------------	-----------------	--------------	----

	μ						
GEV1	0.500	0.750					
GEV2	0.9995	0.9999					



4.2.2. Modelos basados en excedencias sobre un umbral

Debido a la escasez de valores extremos que se pueden llegar a obtener al aplicar los modelos basados en bloques de máximos, puede ocurrir que se terminen requiriendo muestras de datos excesivamente largas para conseguir un mínimo número aceptable de valores extremos. Esto, sumado a la rigidez que conlleva asociar la selección de valores a bloques de tiempo, puede degenerar en que se terminen ignorando valores cuando dentro de un mismo bloque suceda más de un evento que se puede considerar como extremo.

Como alternativa a este tipo de modelos se suele optar por considerar como valores extremos a aquellos que excedan un determinado umbral u. A partir de los valores que exceden dicho umbral se escoge el máximo pico de cada evento identificado. En este ámbito, es interesante denotar que los eventos, para ser considerados como tales, deben cumplir unos estándares de independencia y duración, los cuáles no son sencillos de establecer bajo un criterio riguroso y fundamentado. Desde una perspectiva paramétrica, un evento es comúnmente considerado como una secuencia de estados de mar con valores de la variable que se encuentran por encima de u para un mínimo periodo de tiempo establecido, de manera que cuando la duración del evento es menor a la establecida, este no es considerado. Por otro lado, se asume que dos eventos consecutivos deben ser considerados como uno mismo si los valores de la variable no decaen por debajo de *u* durante un periodo mayor al de un cierto intervalo de tiempo mínimo, denominado comúnmente como criterio de independencia. No obstante, como se ha comentado, esta definición no es completamente rigurosa y libre de ambigüedades que deben ser minimizadas en la medida de lo posible. Para establecer un criterio de independencia adecuado, se debe asegurar que, por un lado, cumpla las bases matemáticas de partida y, por el otro, su duración se base en la física del fenómeno extremo al que esté asociada la variable. Para este trabajo, considerando que en la literatura se recomienda un rango de valores que alberga desde los 1.25 hasta más de 10 días, se ha optado por establecer un criterio de independencia de 72 h. Sin embargo, no existen criterios preestablecidos para seleccionar este parámetro y, como resultado, se dispone de una gran variedad de procedimientos sustancialmente diferentes (Guerra & Rodríguez, 2020; Luceño, 2006; Menéndez, 2008)

En consecuencia, todavía no existe una definición completamente aceptada y en la literatura se han propuesto muchas definiciones alternativas. La mayoría de estas están basadas en la estadística subyacente de la TVE, específicamente en el modelo de excedencias sobre un umbral o, en inglés, Peaks Over Thresold (POT).

4.2.2.1. Modelo GPD – P

Con esta metodología, el teorema de Pickands sostiene que, dada una muestra de valores independientes e idénticamente distribuidos, la distribución de los datos que exceden un cierto umbral converge a una Distribución Generalizada de Pareto (*GPD*) (*Vega et al., 2013*). La GPD será, por tanto, para excedencias sobre un umbral, la distribución análoga a la GEV para máximos sobre un bloque de tiempo. Su CDF es:

$$G(x; \sigma, \xi) = \left\{ 1 - \left(1 + \xi \frac{x}{\sigma} \right)^{-1/\xi} \quad \text{para } \xi \neq 0$$
(12)

mientras que la PDF será:

$$g(x; \sigma, \xi) = \left\{ \frac{1}{\sigma} \left(1 + \frac{x}{\sigma} \right)^{-\frac{1}{\xi} - 1} \quad \text{para } \xi \neq 0$$
 (13)





En este trabajo, se ha realizado la combinación del modelo de Poisson para el número de ocurrencias por unidad de tiempo y el GPD para la intensidad de estas. Para ello se asume que las excedencias para un año dado siguen una distribución de Poisson El modelo que deriva de esta combinación se denomina modelo de Pareto-Poisson y se puede expresar de manera compatible con la distribución GEV para máximos anuales mediante las relaciones:

$$\sigma = \psi + \xi(u - \mu); v = (1 + \xi(u - \mu) / \psi)^{-1/\xi}$$
(14)

La idea fundamental es la de analizar las propiedades estadísticas de las excedencias sobre un umbral en cuanto a la distribución de las intensidades de estas. El teorema asegura que, si se escoge un umbral lo suficientemente elevado y unos parámetros apropiados ($\sigma y \xi$), la distribución de excesos ($F_u(x)$) para un umbral se encuentra dentro del máximo dominio de atracción de G(x) ([$F_u \in MDA(G)$]), por lo que $F_u(x) \approx G(x)$, permitiendo así estimar la cola de la distribución. Sin embargo, es tarea difícil establecer un umbral que conjuntamente satisfaga la independencia de los eventos y permita un número suficiente de eventos para que la muestra sea estadísticamente representativa. Por lo tanto, debido a la incertidumbre asociada a la selección del umbral, existe una gran variabilidad de criterios a la hora de establecer este parámetro, dependiendo principalmente de la localización geográfica y el clima marítimo general (*Di Paola et al.,* 2020). En el subapartado 4.2.2.2.1 se ha añadido una breve discusión en mayor detalle sobre el modelo exploratorio para evaluar la selección del umbral.

Respecto al parámetro de forma, ξ , el teorema demuestra que este es el mismo que en el caso de la GEV, definiendo así el comportamiento de la cola de la distribución y, por tanto, contemplando 3 casos diferentes en función de su signo:

- \succ ξ > 0. Corresponde a la distribución de Pareto y es válida para 0 < x < ∞.
- \succ ξ < 0. Equivale a la familia de Weibull. Válida para 0 < x < ∞.
- > $\xi = 0$. Pertenece a la distribución de Gumbel; se simplifica de la ecuación (X.x) a la que se muestra en el siguiente apartado.



Figura 8. PDF de Pareto para diferentes valores de ξ .

4.2.2.2. Exponencial



UNIVERSIDAD DE CANTABRIA

La familia de distribución exponencial es un caso especial de la familia de GPD, análogo al caso del modelo de Gumbel con la GEV, para la que $\xi = 0$ y tiene media σ . Su CDF es:

$$G(x) = 1 - exp\left(-\frac{x}{\sigma}\right); \quad \text{para } \xi = 0 \tag{15}$$

mientras que la PDF:

$$g(x) = \frac{1}{\sigma} exp\left(-\frac{x}{\sigma}\right); \quad \text{para } \xi = 0$$
 (16)

4.2.2.2.1. Selección del umbral

Si bien a la hora de definir una muestra de eventos extremos se debe considerar el método de extracción, las características de esta selección dependen del modelo de que extremos que se desee aplicar. En el caso del modelo POT, como se ha venido comentando, el método que se emplea es el de excedencias sobre un umbral. Este parámetro es de gran relevancia, ya que, junto con el criterio de independencia descrito anteriormente, condicionará las condiciones que se puedan obtener de la aplicación del modelo.

Con la selección del umbral se busca siempre obtener el mayor número de valores extremos posibles, ya que de esta manera se favorece la robustez del ajuste. Sin embargo, esta práctica debe realizarse con mesura, ya que, considerando las aproximaciones de las que parte modelo estadístico, aplicar un valor excesivamente bajo afecta a la estabilidad del resultado (*Menéndez, 2008*).

Uno, de entre los diferentes métodos exploratorios que existen, consiste en el análisis de los gráficos de la vida de la media residual (*mean residual/excess plot*). Estos se calculan a partir de las medias de las muestras de extremos obtenidas para diversos umbrales (*Davidson & Smith, 1990*). El umbral, con estos gráficos, se identifica a partir del punto en el que la gráfica comienza a perder la continuidad en su tendencia (*figura 9*). En el ejemplo que se muestra, se observa una tendencia lineal hasta que el valor del umbral es excesivamente alto y empieza a generar inestabilidades (línea roja). Aquí se puede ver que la media residual del TWL se vuelve inestable ligeramente por encima del umbral de 3.35 m. Se representan también los intervalos de confianza al 95% de las medias residuales, para lo que se asume normalidad en las muestras de las medias.

Como se puede ver en los paneles **b** y **c** del gráfico, en estos análisis también se puede considerar la opción de añadir los valores que otros indicadores adoptan al modificar el umbral, como son en este caso el periodo de retorno de 50 años (*R50*) o el parámetro lambda que representa el número de eventos registrados por año. En el caso del modelo GPD, también se podría añadir el parámetro de forma (ξ) o, si se aplica el modelo GDP – P, los parámetros equivalentes a la GEV (μ, ψ, ξ).

No obstante, no siempre se pueden observar límites claros a partir de los cuales las muestras se vuelvan inestables, tal y como se puede ver, el R50 sí llega un punto a partir del cual modifica bruscamente su tendencia (coincidiendo prácticamente con el de el parámetro σ) pero para λ este punto no es tan visualmente claro, pudiendo esto dificultar la selección del umbral en ciertos casos.

Sin embargo, según *Bernardara et al. (2014)*, a pesar de las distintas metodologías existentes, algunos autores establecen el umbral usando un cuantil concreto de la serie temporal de la variable de estudio. *Ruggiero et al. (2010)* lo establece en el percentil 99.5 para variables oceánicas como la altura de ola. Mientras que otros como *Rosbjerg et al. (1992)* sugieren calcular el umbral físico como el valor medio de la serie observada sumada tres veces la desviación estándar.







Figura 9. Gráfico de vida residual con intervalos de confianza al 95%.

En este trabajo, se ha decidido aplicar el primer criterio para seleccionar el umbral a la hora de ajustar las muestras de extremos con los modelos GPD y Exponencial. Por un lado, se ha establecido como umbral el cuantil 99.5 para las variables cuyo forzamiento es de origen atmosférico (NTR y η), mientras que, para el TWL, después de realizar varias pruebas, se consideró que era más adecuado aplicar el cuantil 99.7. Esto es debido a que, al tener gran parte de contribución armónica la AT en su valor final, la media de la muestra de extremos se ve mucho más condicionada por los valores más bajos de la misma, ya que son más abundantes.

4.3. Significancia de ξ

Para analizar si los modelos GEV y GPD realizan un mejor ajuste que los modelos con $\xi = 0$, es decir, Gumbel y Exponencial respectivamente, se analiza la significancia de ξ . De esta manera, se puede saber en qué localizaciones se puede considerar estadísticamente significativa la existencia o no de distintas familias de extremos.

Para el caso del modelo GEV, esto se ha realizado aplicando el método denominado como prueba de la razón de verosimilitud, con el cual se estudia el parámetro de la desvianza (D_s). El procedimiento consiste en obtener D_s comparando los valores estimados de las funciones logarítmicas de verosimilitud de los modelos que difieren en el número de parámetros incluidos ([μ , ψ , ζ] y [μ , ψ]). Si se define la primera función logarítmica: $l(x; \theta^a)$ para el primer modelo, donde θ^a es un vector cuya dimensión p son el número de parámetros de la distribución, y $l(x; \theta^b)$, para el segundo modelo con un número de parámetros p – 1, siendo 1 el número de grados de libertad (q); D_s vendrá dada por:

$$D_S = 2 \cdot \left[(l(x; \theta^a) - l(x; \theta^b)) \right]$$
(17)

A partir de las propiedades del método de máxima verosimilitud se puede deducir que $D_S \approx X_q^2$, donde X_q^2 es la función chi-cuadrado con q grados de libertad. Por lo que, considerando la probabilidad 1 - α , se puede





asumir que la mejora de los q parámetros es significativa al nivel α si D_s es mayor que el valor $X^{2}_{q,1-\alpha}$ (*Menéndez*, 2008).

Por otro lado, para el modelo basado en el método POT, se han calculado las bandas de confianza del ξ de cada ajuste. En este caso, dicho parámetro será significativo cuando el rango entre el intervalo inferior y superior no pase por cero.

5. Resultados y Discusión

5.1. Contribuciones al TWL

Realizado el análisis armónico, se obtuvieron los registros de la variable NTR, a la par que se obtuvieron los de η y TWL tras los cálculos aplicados. Con esto, y una vez obtenidas las muestras de extremos, se procedió a calcular las contribuciones medias (en tanto por uno) del NTR (*NTRcon*) y del η (*Setupcon*) a los máximos anuales de TWL. Esto se hizo con el fin de representar y distinguir las zonas presentan una mayor influencia de la componente meteorológica en los eventos extremos. En la *Figura* se representan las contribuciones medias para cada punto. En el panel (**a**), las zonas con mayor aporte del NTR son las que engloban la zona norte del Golfo de México, este de Norteamérica, el Mar del Norte y norte de Australia, con valores por encima de alrededor del 20% del TWL. En cambio, las zonas en las que se aprecia una menor contribución de la componente meteorológica del nivel del mar son el sur de Oceanía y polinesia (17, 16 y 18 respectivamente), junto con la costa oeste de Norteamérica (20 y 21) y, con diferencia, la zona este de África (19), con aportes menores a alrededor del 10%. Por lo general, es posible observar una contribución del NTR considerable, aunque menor que la del η , llegando en algunos casos a ser muy débil e incluso negativa (19 y 21). En línea con lo que se citaba al inicio del documento, la contribución al cambio del nivel del mar debida a los descensos de la presión atmosférica es significativamente menor que la causada por la acumulación de agua contra la costa por la acción del viento (*Pugh, 1987*).

En el panel (**b**), se puede observar que en islas como Reunión y la Polinesia, en la costa del Pacífico y en el Golfo de México se encuentran las mayores contribuciones del η a los eventos extremos, llegando a suponer un 80% del valor de TWL en el norte de Australia y este de África. Son menores, aunque importantes, en ciertos puntos de las costas del Atlántico, con alrededor de un 30% de aporte, y despreciables en el Mar del Norte (<5%). También son considerables en el norte del Golfo de México (20% - 30%), mientras que más al sur, en la zona del Caribe (punto 4), llegan a superar el 50%. Esto parece coherente ya que dicho punto está en Puerto Rico; una isla que está menos protegida del océano abierto y, por tanto, más expuesta a mayores vientos, que generan mayores acumulaciones de masas de agua en la costa. Algo similar se podría decir que sucede en La Reunión (19).

Por lo general, se puede ver que el η suele ser mayor en las islas oceánicas, expuestas a swells de largo periodo. Además, los márgenes oceánicos con prolongada plataforma continental favorecen mayores valores de NTR en comparación a las islas, donde la plataforma suele ser de menor longitud (*Pinet, 2003*).







Figura 10. Contribución media del NTR (a) y η (b) a los valores extremos del TWL para los 21 puntos de estudio.

Tanto estos parámetros como otros adicionales se resumen en la tabla (XX) que se encuentra en el Anexo.

5.2. Optimización del ajuste mixto







Figura 11. Ajuste de las distribución mixta a la variable TWL antes y después de aplicar la optimización.

A continuación, previo a la presentación de los resultados principales, se ha añadido la *figura 11* como breve ejemplo ilustrativo de la metodología aplicada para la optimización de los ajustes de las distribuciones mixtas. En rojo muestra el ajuste realizado con los límites del parámetro μ que se pretendía establecer inicialmente ($\mu_{GEV1} >= 0.75$; $\mu_{GEV2} >= 0.9995$) tal y como se muestra en la leyenda, mientras que en verde se representa el resultado para el que se encontró el mejor ajuste con la función de optimización. Como se puede ver, se consigue una mejora en los ajustes a los valores empíricos, describiendo de manera más precisa la trayectoria que describen los datos. Además, también se observa una mejora en su predicción, mucho más realista respecto a los valores esperables para periodos de retorno de 100 años. Mientras que con el primer ajuste se devuelven valores de 3.9 m, después de aplicar la optimización se observa cómo se consigue controlar la curva y reducirlos a valores más similares a los observados (3.6 m).

Para añadir, también es remarcable la diferencia en los parámetros de localización de los dos ajustes; si comparamos la función GEV2 de ambos ajustes, se consigue un mejor resultado con la que se establece el límite en el cuantil más alto (0.9995). Se puede interpretar que esto de se deba a que esta modela mejor los valores más extremos al restringir la aplicación de la función en la zona más alta de la distribución de la muestra. Por otro lado, si bien el μ correspondiente a la segunda función GEV sí se modifica, el que está asociado a la primera no varía entre ambos ajustes, esto denota la mayor sensibilidad asociada al parámetro de la segunda GEV que al de la primera, debido a que se está ajustando a la parte más alta de la distribución de la muestra de extremos. Esto se refleja en la mayor similitud de los ajustes que se observa para periodos de retorno pequeños al compararla con la que se visualiza para los más altos.

5.3. Evaluación de la significancia

Los resultados obtenidos para evaluar si ξ es significativo se muestran en la tabla XX. Según la teoría de base de este análisis, que ξ sea significativamente distinto de cero, es un indicador de que la distribución Gumbel (o Exponencial) no es adecuada para modelar cierta variable en una determinada localización. Esto debido a que la tendencia dominante en los datos se aleja de la de una recta. Adicionalmente, se puede





extraer que, si el parámetro ξ es significativo y negativo, se podría interpretar que los valores siguen un comportamiento común y básicamente obedecen a un mismo forzamiento. Mientras que, si es significativo y positivo, puede que estos tengan su génesis en fenómenos distintos y, por tanto, exista más de una familia de valores extremos en la muestra que sigan comportamientos distintos.

		NTR			TWL				η				
		G	EV	GF	PD	G	EV	G	PD	G	EV	G	PD
1	St Petersburgo, USA (Florida)	1	+	0	+	1	+	1	+	1	-	0	+
2	Pensacola, USA (Alabama)	1	+	0	+	1	+	1	+	1	+	0	+
3	Galveston, USA (Texas)	1	+	0	+	1	+	1	+	1	+	1	+
4	Magueyes, Puerto Rico	1	+	1	+	1	+	1	+	1	+	0	+
5	Montauk, USA (Nueva York)	1	-	0	-	1	+	0	+	1	+	0	+
6	Lewes, USA (Delaware)	1	-	0	-	1	+	1	+	0	-	0	+
7	Cape May, USA (Delaware)	1	-	0	-	1	+	0	+	1	-	0	-
8	New London, USA (Connectica)	0	-	0	+	1	+	0	+	1	+	0	+
9	Cuxhaven, Alemania	1	-	1	-	1	-	0	-	1	+	0	+
10	Naha, Japón	1	-	0	+	1	-	0	-	1	-	1	-
11	Miyakejima, Japán	1	-	0	+	1	+	0	+	0	+	0	+
12	Wakkanai, Japón	1	+	0	+	1	-	0	-	1	-	0	-
13	Cananeia, Brasil	1	+	0	+	1	-	0	-	1	-	1	-
14	Apra, Guam	1	-	1	+	1	-	0	+	1	+	0	+
15	Weipa, Australia	1	-	0	-	1	-	1	-	1	-	1	-
16	Pago Ago, USA	1	+	1	+	1	-	1	-	1	+	0	+
17	Wellington, NZ	1	-	1	-	1	+	0	+	1	+	0	+
18	Polinesia francesa, Francia	1	+	1	+	1	-	1	-	1	-	0	-
19	Pointe de Galets, Reunión	1	-	0	+	1	-	0	-	1	-	0	-
20	Yakutat, USA	1	-	1	-	1	-	1	-	1	-	0	-
21	San Francisco, USA	1	-	0	-	1	-	1	-	1	-	0	-

Tabla 2. Resultados de la evaluación de la significancia del parámetro ξ para cada punto junto con su signo. En rojo los que sí son significativamente distintos de 0; en blanco los que no.

Si observamos la tabla, se puede ver cómo, para el modelo GEV, no es significativo sólo en 3 de los 63 ajustes. Por el contrario, cuando se realiza el ajuste con Gumbel, se observa un mayor número de casos en los que esto no se cumple, concretamente en 40 de los 63 ajustes. Esto podría deberse a que, al obtener muchos más valores para las muestras de extremos con el método POT que con el método de selección por máximos anuales, los valores de la parte alta de la cola tiendan a no distinguirse en igual medida de los que tienen pertenecen a la zona baja de la distribución de la muestra, repercutiendo esto en el análisis realizado con la función X^2_q . Esto podría ser un indicador de que el umbral escogido para la variable no es el más adecuado para en determinado punto de estudio. Se podría intentar mejorar el análisis modificando el umbral en los casos en los que el resultado no coincidiese entre los dos modelos.

Los tres casos en los que se obtiene que el parámetro no es significativo con el modelo GEV corresponden a la variable NTR, en el punto 8 (Este de Norteamérica), y al η en los puntos 6 y 11 (Este de Norteamérica y Asia). Por lo que se podría interpretar que, en estos casos, las muestras de extremos se pueden modelar con la distribución de Gumbel, como se muestra en los ajustes de ejemplo de la *figura 12*.

Por otro lado, de los casos en los que se obtiene un resultado significativo y positivo en ambos, GEV y POT, destaca la zona del Golfo de México y Caribe para las variables NTR y el TWL. Para η sólo se obtiene con GEV, exceptuando el punto 3 (Texas), que presenta este patrón en todas las variables. Esto podría considerarse un indicador que denota influencia que los ciclones tropicales en dichas zonas.







Figura 12. Ajustes de GEV y Gumbel con ξ no significativo para los puntos 8, 6 y 11 respectivamente.

5.4. Comparativa y análisis de los diferentes modelos

Una vez ya realizados los ajustes, en la *figura 13* se muestran algunos en gráficos de periodo de retorno. Se representan las tres variables en algunos de los puntos de estudio (17, 15, 4, 18, 14 y 13). Por columnas:





modelos GEV y Gumbel; modelos GPD y Exponencial; y mixtas. Por filas: TWL (filas 1 y 2); NTR (2 y 3) y η (4 y 5). Con estos gráficos se muestra la cola de la distribución en un sola representación con la finalidad de que los niveles de retorno asociados a periodos de retorno altos se puedan apreciar con facilidad. Al representar los distintos modelos, la linealidad de la representación cuando $\xi = 0$, tanto para el método de máximos anuales como para el método POT, proporciona una línea base respecto a la que juzgar el efecto del ξ estimado (Coles, 2001). Adicionalmente, a cada gráfico se le ha añadido los valores de las muestras de extremos obtenidas con sus respectivos métodos de selección (máximos en un bloque de tiempo y excedencias sobre un umbral). En este sentido, cabe comentar que, si bien las funciones GEV y las mixtas se basan en el mismo método de selección, se han aplicado criterios distintos para establecer los bloques temporales, siendo este anual para las primeras y semanal paras las mixtas. Esto porque a la hora de realizar los ajustes de las segundas, basado en la bibliografía, se ha observado que es el bloque temporal más adecuado. Por otro lado, se ha añadido a cada modelo sus correspondientes intervalos dentro de los cuales se estima que se encontrarán los valores empíricos respecto a los ajustados con un nivel de confianza del 95 %.

Al comparar los modelos entre sí, por un lado, se puede ver que los modelos clásicos que siguen distintos métodos de selección describen un comportamiento razonablemente similar. Si bien en la primera fila el modelo GPD infraestima los valores más altos de la distribución (periodos de retorno superiores a 10 años), el modelo GEV se acerca bastante más a estos.

Para las filas 2 y 6 (puntos 15 y 13) se observa que $\xi < 0$ y que, por tanto, la distribución de los datos de las variables pertenece a una familia de tipo Weibull para esas variables (TWL / η), siendo entonces las funciones GEV y GPD las idóneas de entre las distribuciones clásicas para describir el comportamiento de estos extremos. No obstante, al mostrar el primer punto un valor cercano a cero con el modelo GEV, se podría cuestionar la significancia del parámetro ξ y, en consecuencia, si con el modelo Gumbel fuese suficiente para su correcto modelado. Sin embargo, si se revisa la significancia a partir de lo obtenido en la tabla XX, se puede confirmar que este es significativamente distinto de cero en ambos casos.

Respecto a las bandas de confianza, con la función exponencial tienden a ser más estrechas que las de la distribución GPD, excluyendo los valores más extremos de su rango. Esto también sucede de forma similar con las funciones GEV y Gumbel, aunque en con la función de Gumbel sí que se llegan a abarcar los valores más extremos en algunos casos. Esta diferencia entre las funciones Gumbel y Exponencial puede deberse al umbral establecido durante la selección de la muestra, ya que, si este no es lo suficientemente elevado, la parte baja de la distribución de la muestra condicionará la tendencia que marque la función Exponencial, excluyendo de los intervalos de confianza los valores más extremos. En línea con esto, las dos últimas filas, que muestran un mejor ajuste de la Exponencial (ya que excluyen menos valores del rango esperado), son las que presentan un menor número esperado de eventos por año (λ), el cuál es un parámetro directamente relacionado con el valor del umbral; si λ es bajo significa que se ha utilizado como umbral un cuantil elevado. Sin embargo, es relevante comentar que estos dos casos son bastante cercanos a uno (1.21, 1.58), lo cuál de por sí es un valor considerablemente bajo teniendo en cuenta que el sentido del método POT reside en obtener más valores extremos para las muestras que con el método de selección por máximos anuales, con el que se registra un evento por año ($\lambda = 1$).

Si se comprara la calidad del ajustes previos con los de las distribuciones mixtas, se puede ver que el comportamiento en el caso de las segundas es considerablemente mejor. El desempeño de este método depende de las componentes de mezcla que se escojan, sus respectivos pesos y los valores de sus parámetros, como se comentaba en la descripción de la metodología. Para este caso concretamente, parece que se obtiene un resultado fiable al aplicar una distribución normal para las zonas medias y bajas de la







Figura 13. Ajustes de los 5 modelos por separado para los puntos 17, 15, 4, 18, 14 y 13 respectivamente.





distribución y dos distribuciones GEV para las partes medio alta y alta. Si se compara entre distintas localizaciones para las 3 variables, se observa que siempre ofrece un ajuste preciso. Para la variable NTR, describe mejor la trayectoria que siguen los valores empíricos, tanto cuando todos siguen la misma tendencia (15, Norte de Australia) como cuando presentan más de una (17, Sur de Oceanía). Lo mismo sucede con la variable TWL, realizando incluso predicciones más moderadas para periodos de retorno más altos (100 años). Sin embargo, para la variable η del punto 14 (Este de Asia), sí realiza predicciones más elevadas que los otros modelos; estima valores de hasta 1 m por encima que el resto y no llega a estabilizar la curva para los periodos de retorno más altos. Por lo que no se puede decir, por lo menos con las componentes de mezcla seleccionadas, que siempre realice predicciones más acertadas. En resumen, las distribuciones mixtas presentan ajustes similares en trayectoria a los modelos GEV y GPD pero con mejores predicciones para los periodos de retorno más altos.

Como parte de resultados del análisis, en la *figura 14* se representan los ajustes de los cinco modelos para las tres variables en dos puntos distintos cada una. Cada fila representa una variable y la columna izquierda zonas de ciclones extra-tropicales mientras que la derecha de ciclones tropicales. También se han añadido los valores empíricos de las muestras obtenidas con el método de máximos anuales para resaltar la calidad de los distintos ajustes. Los paneles de la izquierda corresponden a los puntos 13, 21 y 9 (Este Sudamérica, Oeste Norteamérica y Mar del Norte), mientras que los de la derecha pertenecen a los puntos 2, 17 y 7 (Golfo de México, Sur de Oceanía y Este de Norteamérica). En primer lugar, como distintivo principal, se puede ver cómo los ajustes clásicos en zonas de ciclones tropicales se aproximan a una distribución de Frechet ($\xi > 0$), ya que, exceptuando el panel f, se observan comportamientos que parecen asintóticos, al ver que las predicciones que realizan (GEV y GDP) para periodos de retorno altos se disparan y no son realistas desde el punto de vista físico de la variable, incluso sobreestimando en algunos casos los valores para los periodos de retorno más altos (panel b y d). Esto es esperable ya que, como en otros estudios se comenta, en regiones donde los SS más grandes están controlados por ciclones tropicales, la diferencia media entre los niveles extremos del mar empíricos y ajustados para periodos de retorno altos es significativamente mayor que la que se encuentra en regiones extra-tropicales (*Muis et al. 2016*).

En el caso de las zonas con tormentas extra-tropicales, se ve que estas no presentan unos ajustes tan pronunciados en comparación, ya que como se ha mencionado, los valores extremos obedecen a unos mismos fenómenos atmosféricos. No obstante, esto no siempre se cumple, ya que, por ejemplo, el panel e, que corresponde a registros de η en el Mar del Norte, muestra una tendencia bastante acusada con el modelo GEV.

Si nos centramos en la calidad de los ajustes, los modelos clásicos describen con considerable precisión para periodos de retorno bajos y medios; tanto para el TWL en el panel d como para el η en f, los modelos Gumbel y Exponencial ($\xi = 0$) sirven para ajustar incluso a periodos de retorno de 25 y 50 años sin excesivo error. Sin embargo, para variables como el NTR no parecen ser tan adecuados. Una mejora significativa parece que ofrecen los modelos GEV y GPD para los periodos de retorno más altos ya que, si bien también presentan errores considerables en las estimaciones, describen mejor la curvatura que adoptan los valores empíricos en los casos en los que la presencia de más de una familia es evidente. En último lugar, se puede ver con relativa claridad cómo las funciones mixtas (línea azul oscuro) son las que mejor describen el comportamiento de las variables, ya que ni sobreestiman ni infraestiman en igual medida. Por otro lado, estas describen con mayor precisión la curvatura que los valores adquieren cuando esto cambian de su tendencia, pasando de una familia de extremos a otra. Adicionalmente, al contrario que los modelos GEV y GPD, en todos los paneles se observa cómo sus predicciones más allá de los empíricos no se disparan a valores que no sigan una lógica desde el punto de vista de los forzamientos, es decir, que no estén delimitadas por las condiciones físicas del entorno. En resumen, la comparación de los valores de retorno







Figura 14. Ajustes de los 5 modelos para las 3 variables en 2 puntos diferentes para cada una. **a**. Cananeia, (13); **b**. Pensacola, (2); **c**. La Jolla, 28; **d**. Wellington, 17; **e**. Cuxhaven, 9; **f**. Cape May, 7.

empíricos con los ajustados por este modelo muestra resultados más robustos. Quedando claro con esta representación la habilidad de modelar tanto los valores más habituales como los extremos más inusuales, incluso cuando se observa que los valores máximos revelan claras diferencias en su comportamiento en la parte más alta de la cola.



5.5. Parámetros de las distribuciones

En las tablas **3**, **4** y **5** se muestran los parámetros de todos los ajustes realizados para cada localización, descritos a partir del apartado **4.2.1**. Para facilitar su interpretación, se han añadido distintas clasificaciones por rangos de colores a los parámetros a lo largo de las columnas, con el fin de ayudar a su lectura.

En primer lugar, se puede ver que las medias son considerablemente mayores para las muestras anuales que las obtenidas a partir de un umbral. Esto porque las primeras se generan sólo con el valor más alto de cada año, mientras que para las segundas se recogen muchos más valores que, en su conjunto, aunque dependan de lo elevado que sea el umbral, tenderán a ser menores porque siempre se intenta que dicho parámetro sea lo más bajo posible para conseguir una mayor cantidad de valores.

Tablas 3, 4 y 5. Parámetros de las distribuciones de las 3 variables para los 21 puntos de estudio. Para el ξ en rojo los valores más altos y verde los más bajos. Para μ los más en rojo y los más bajos en azul. Para lambda los más altos en verde intenso y más bajos en amarillo.

			NTR								
			GEV		Gur	nbel	Gl	PD	Expon	encial	
		μ	ψ	ξ	μ	ψ	σ	ξ	σ	λ	
1	St Petersburgo, USA (Florida)	0.585	0.209	0.028	0.588	0.211	0.135	0.228	0.172	3.095	
2	Pensacola, USA (Alabama)	0.468	0.163	0.520	0.521	0.229	0.174	0.272	0.233	1.634	
3	Galveston, USA (Texas)	0.598	0.222	0.084	0.609	0.228	0.100	0.384	0.159	2.051	
4	Magueyes, Puerto Rico	0.164	0.056	0.228	0.171	0.063	0.025	0.335	0.038	2.244	
5	Montauk, USA (Nueva York)	0.705	0.219	-0.079	0.695	0.217	0.184	0.006	0.185	3.610	
6	Lewes, USA (Delaware)	0.910	0.222	-0.208	0.885	0.214	0.223	-0.101	0.203	3.190	
7	Cape May, USA (Delaware)	0.812	0.212	-0.131	0.797	0.206	0.176	-0.012	0.174	3.190	
8	New London, USA (Connectica)	0.741	0.215	0.011	0.743	0.216	0.161	0.111	0.181	4.548	
9	Cuxhaven, Alemania	2.538	0.491	-0.251	2.475	0.455	0.719	-0.217	0.587	4.231	
10	Naha, Japón	0.438	0.150	-0.076	0.432	0.146	0.111	0.158	0.130	1.900	
11	Miyakejima, Japán	0.521	0.248	-0.042	0.515	0.244	0.066	0.397	0.103	1.500	
12	Wakkanai, Japón	0.405	0.077	0.191	0.413	0.084	0.072	0.085	0.078	3.325	
13	Cananeia, Brasil	0.766	0.094	0.231	0.779	0.106	0.115	0.051	0.122	4.500	
14	Apra, Guam	0.196	0.081	-0.017	0.195	0.081	0.025	0.204	0.031	2.857	
15	Weipa, Australia	1.035	0.200	-0.134	1.021	0.192	0.270	-0.237	0.217	1.675	
16	Pago Ago, USA	0.108	0.101	0.068	0.112	0.103	0.023	0.223	0.030	2.077	
17	Wellington, NZ	0.372	0.095	-0.264	0.360	0.087	0.077	-0.188	0.064	3.333	
18	Polinesia francesa, Francia	0.156	0.064	0.119	0.160	0.067	0.024	0.330	0.037	2.744	
19	Pointe de Galets, Reunión	0.310	0.112	-0.161	0.300	0.107	0.066	0.084	0.072	1.829	
20	Yakutat, USA	0.648	0.217	-0.470	0.597	0.196	0.163	-0.362	0.118	2.333	
21	San Francisco, USA	0.217	0.067	-0.045	0.215	0.067	0.070	-0.253	0.056	1.538	





			-	-		TWL	-	-			
			GEV		Gur	nbel	Gl	SPD Ex		Exponencial	
		μ	Ψ	ξ	μ	Ψ	σ	ζ	σ	λ	
1	St Petersburgo, USA (Florida)	2.608	0.196	0.172	2.627	0.212	0.165	0.155	0.194	2.952	
2	Pensacola, USA (Alabama)	4.595	0.375	0.517	4.715	0.551	0.258	0.652	0.546	1.537	
3	Galveston, USA (Texas)	3.002	0.197	0.409	3.053	0.259	0.141	0.408	0.232	1.667	
4	Magueyes, Puerto Rico	2.879	0.237	0.237	2.911	0.268	0.187	0.281	0.255	1.610	
5	Montauk, USA (Nueva York)	4.120	0.449	0.070	4.138	0.461	0.383	0.052	0.404	2.951	
6	Lewes, USA (Delaware)	4.110	0.369	0.032	4.117	0.374	0.244	0.237	0.315	3.678	
7	Cape May, USA (Delaware)	3.557	0.178	0.057	3.563	0.182	0.144	0.075	0.156	5.299	
8	New London, USA (Connectica)	3.864	0.342	0.074	3.878	0.350	0.272	0.091	0.300	4.355	
9	Cuxhaven, Alemania	8.615	0.435	-0.054	8.602	0.426	0.537	-0.121	0.478	4.744	
10	Naha, Japón	6.080	0.730	-0.102	6.041	0.701	0.774	-0.096	0.705	2.225	
11	Miyakejima, Japán	6.410	0.785	0.077	6.443	0.813	0.661	0.069	0.709	2.550	
12	Wakkanai, Japón	4.296	0.312	-0.019	4.293	0.310	0.276	0.067	0.296	2.850	
13	Cananeia, Brasil	3.727	0.144	-0.153	3.715	0.139	0.142	-0.044	0.136	5.173	
14	Apra, Guam	3.444	0.547	0.036	3.455	0.554	0.429	0.143	0.498	1.524	
15	Weipa, Australia	3.844	0.211	-0.082	3.835	0.206	0.239	-0.134	0.211	2.150	
16	Pago Ago, USA	2.038	0.104	-0.174	2.029	0.100	0.096	-0.342	0.070	3.103	
17	Wellington, NZ	3.964	0.192	0.388	4.011	0.246	0.179	0.230	0.233	2.924	
18	Polinesia francesa, Francia	3.688	0.248	-0.125	3.672	0.244	0.254	-0.129	0.226	2.462	
19	Pointe de Galets, Reunión	2.964	0.198	-0.130	2.951	0.197	0.191	-0.035	0.184	1.658	
20	Yakutat, USA	7.539	0.602	-0.539	7.383	0.560	0.658	-0.330	0.492	4.137	
21	San Francisco, USA	6.703	0.419	-0.088	6.683	0.409	0.459	-0.090	0.421	4.205	

			η							
			GEV		Gur	nbel	G	PD	Exponencial	
		μ	ψ	ζ	μ	ψ	σ	ξ	σ	λ
1	St Petersburgo, USA (Florida)	0.580	0.096	-0.247	0.568	0.088	0.071	0.106	0.079	3.548
2	Pensacola, USA (Alabama)	1.328	0.232	0.668	1.429	0.368	0.293	0.243	0.380	1.805
3	Galveston, USA (Texas)	0.848	0.076	0.659	0.881	0.123	0.033	0.748	0.085	4.256
4	Magueyes, Puerto Rico	1.676	0.228	0.117	1.691	0.239	0.199	0.153	0.235	2.488
5	Montauk, USA (Nueva York)	1.894	0.308	0.238	1.935	0.347	0.275	0.108	0.309	3.341
6	Lewes, USA (Delaware)	1.434	0.222	-0.004	1.434	0.222	0.196	0.028	0.202	3.595
7	Cape May, USA (Delaware)	0.887	0.110	-0.019	0.886	0.109	0.108	-0.015	0.106	4.143
8	New London, USA (Connectica)	1.544	0.213	0.154	1.562	0.228	0.197	0.084	0.215	4.000
9	Cuxhaven, Alemania	0.363	0.065	0.336	0.376	0.079	0.068	0.131	0.078	3.308
10	Naha, Japón	2.729	0.678	-0.092	2.697	0.655	0.839	-0.170	0.713	2.300
11	Miyakejima, Japán	3.438	0.686	0.002	3.439	0.686	0.482	0.185	0.586	3.375
12	Wakkanai, Japón	2.331	0.288	-0.124	2.312	0.282	0.308	-0.065	0.289	3.825
13	Cananeia, Brasil	1.159	0.088	-0.229	1.149	0.082	0.132	-0.347	0.098	2.462
14	Apra, Guam	2.352	0.548	0.049	2.367	0.559	0.461	0.092	0.507	1.667
15	Weipa, Australia	0.887	0.124	-0.291	0.868	0.116	0.130	-0.096	0.118	3.850
16	Pago Ago, USA	0.205	0.026	0.052	0.206	0.027	0.025	0.111	0.028	2.205
17	Wellington, NZ	2.426	0.243	0.293	2.469	0.288	0.252	0.169	0.304	1.821
18	Polinesia francesa, Francia	2.493	0.247	-0.120	2.477	0.241	0.272	-0.138	0.239	3.436
19	Pointe de Galets, Reunión	2.281	0.211	-0.151	2.264	0.215	0.221	-0.093	0.202	1.585
20	Yakutat, USA	4.037	0.521	-0.155	3.995	0.500	0.519	-0.145	0.453	2.500
21	San Francisco, USA	3.945	0.380	-0.064	3.932	0.377	0.405	-0.081	0.375	4.051

Respecto al parámetro de forma, el cual es el principal a analizar, al comparar entre los distintos puntos se observa que es mayor por lo general en la zona del Golfo de México, excepto en 1 de los 4 puntos (puntos del 1 a 4). Este es de partida un resultado esperable, ya que un mayor valor en estos puntos respecto al resto puede ser un reflejo de la influencia de los ciclones tropicales en dicha zona, bien conocidos por su intensidad. Tal y como se podía ver en la *tabla 2*, en la que se observa que esta es la zona con mayores contribuciones del NTR al valor del TWL. Por contrapunto, un punto como el 11 (localizado en el Mar del Norte), a pesar de verse que el NTR presenta también una importante contribución a los valores extremos





de TWL, se puede observar un valor bastante bajo del parámetro de forma (-0.251), denotando esto que los extremos de NTR en este punto siguen una distribución de Weibull, la cual se da cuando las condiciones de contorno explican un lógico valor máximo a partir de un límite físico. Esto es esperable al ser esta una zona cuya componente meteorológica está únicamente dominada por ciclones extra-tropicales.

Por otro lado, en otras zonas que también se sabe que sufren los impactos de ciclones tropicales, pero en menor medida, dicho parámetro no alcanza valores tan elevados. Puntos como el 17 (Nueva Zelanda), presentan algunos eventos extremos que acentúan el valor de ξ para el η , pero no para el NTR. Esto puede indicar que en la zona existen distintos forzamientos de vientos que generen distintos valores extremos de η , pero no de presión atmosférica que generen distintas familias de extremos de NTR.

Por último, en el oeste de Norteamérica (20 y 21), se observa con relativa claridad que para ninguna de las variables existan forzamientos atmosféricos diferentes que generen familias de extremos distintas.

5.6. Modelos óptimos

Como última parte de este análisis, para poder discernir cuál era el modelo que mejores resultados arroja para cada variable en cada punto, se ha optado por comparar el desempeño que estos presentan en términos del error en el ajuste. Para ello se ha calculado el Error Cuadrático Medio Relativo (RRMSE) entre los valores empíricos y los valores ajustados por los modelos en los periodos de retorno de los empíricos. En este sentido, es importante aclarar de antemano que, debido a que los distintos modelos están basados en diferentes métodos para la selección de valores extremos, se debe aplicar esta métrica sobre los valores que estiman sobre unos mismos periodos de retorno para poder así realizar comparaciones entre estos. Esto se hace para evitar que los valores localizados en las partes bajas de las distribuciones de las muestras obtenidas con el método POT, que por lo general son más abundantes, tuviesen un peso excesivo y distorsionasen el resultado final. En este caso se ha optado por realizar el cálculo respecto a los 10 últimos valores de las muestras, permitiendo de esta manera hacer comparables las evaluaciones del parámetro entre distintos modelos, ya que de por sí estos se ajustan a muestras con diferente número de valores. Sin embargo, esta condición puede conllevar su error si el número de valores sobre los que se calcula es demasiado reducido, lo cuál debería estudiarse en futuras investigaciones.

En la *tabla 6* se han añadido los resultados obtenidos para la variable NTR, mientras que los de las variables TWL y η se han ubicado en el Anexo junto con demás figuras para no entorpecer la lectura del documento. En primer lugar, si comparamos los grupos de modelos por método de selección de extremos, vemos que de forma general los que aplican el método POT devuelven valores considerablemente mayores de RRMSE que los que aplican el método de selección de máximos anuales. Puede que el motivo por el que se observan resultados tan pobres para estos modelos sea lo comentado en el párrafo anterior, teniendo que seleccionar más valores para que su error disminuyese. De la misma manera que otras partes del análisis, puede ser que estableciendo un umbral óptimo a las muestras del método POT se consiguiese una mejora en la calidad de los ajustes. No obstante, este resultado tan pobre denota la menor precisión de estos modelos a la hora de ajustar los valores más altos de la distribución.





DMSE			NTR		
RWISE	GEV	Gumbel	Mixtas	GPD	Exponencial
St Petersburgo, USA (Florida)	0.075	0.074	0.043	0.113	0.097
Pensacola, USA (Alabama)	0.209	0.213	0.035	0.269	0.177
Galveston, USA (Texas)	0.276	0.306	0.025	0.351	0.453
Magueyes, Puerto Rico	0.062	0.107	0.012	0.159	0.208
Montauk, USA (Nueva York)	0.107	0.093	0.022	0.152	0.137
Lewes, USA (Delaware)	0.042	0.050	0.022	0.171	0.057
Cape May, USA (Delaware)	0.057	0.038	0.006	0.076	0.087
New London, USA (Connectica)	0.091	0.095	0.011	0.112	0.142
Cuxhaven, Alemania	0.041	0.076	0.059	0.352	0.277
Naha, Japón	0.056	0.053	0.022	0.235	0.122
Miyakejima, Japán	0.053	0.042	0.020	0.092	0.169
Wakkanai, Japón	0.076	0.117	0.012	0.111	0.184
Apra, Guam	0.126	0.158	0.019	0.162	0.222
Cananeia, Brasil	0.078	0.076	0.004	0.123	0.221
Weipa, Australia	0.035	0.027	0.012	0.201	0.059
Pago Ago, USA	0.072	0.091	0.008	0.142	0.219
Wellington, NZ	0.026	0.048	0.004	0.081	0.070
Polinesia francesa, Francia	0.076	0.103	0.010	0.272	0.220
Pointe de Galets, Reunión	0.044	0.034	0.016	0.033	0.083
Yakutat, USA	0.027	0.114	0.008	0.156	0.083
San Francisco, USA	0.030	0.023	0.003	0.087	0.024

Tabla 6. RRMSE obtenido del ajuste de los empíricos de NTR con cada modelo para las 21 localizaciones.

Por otro lado, de entre todos los modelos destacan las mixtas. Se observa que, por lo general, presentan errores considerablemente más bajos que el resto. Adicionalmente, en los puntos en los que las mixtas son el modelo óptimo, los valores de RRMSE suelen ser notablemente menores que en los puntos en los que es otro modelo el que se ajusta mejor, llegando a ser así hasta en 13 de 20 veces para la variable NTR. Esto refleja el nivel de precisión en comparación que puede llegar a presentar este modelo en la predicción de valores asociados a periodos de retorno, el cual, como se observa en los gráficos, es bastante elevado para numerosos casos.

Continuando con las distribuciones mixtas, este es el modelo que mejor se ajusta para las tres variables en el mayor número de localizaciones. Siendo así en 20 de las 21 para la variable NTR; 13 de 21 para el TWL y 11 de 21 para el η . Para el NTR concretamente, este resultado destaca lo bien que modela dicha variable tanto en zonas con influencia de ciclones tropicales como extra-tropicales. En otras palabras, este es un modelo estadístico que, según estos resultados, modela con un desempeño notable tanto cuando existe una como más de una familia de extremos en una muestra.

Finalmente, seleccionados todos los modelos óptimos para cada variable, la comparación resultante para cada punto se muestra en la *figura 15*. Se muestran tres paneles (**a**, **b** y **c**) para las variables NTR, TWL y η respectivamente. Si bien queda claro que las distribuciones mixtas son las óptimas para modelar la componente meteorológica del nivel del mar en cualquier zona, para las variables TWL y η , aun siendo la principal, son más las zonas en las que es un modelo clásico el que devuelve un valor más bajo en el error.

Para el TWL, en el Golfo de México el modelo de funciones mixtas es el que mejor se ajusta en 3 de los 4 puntos, mientras que para el 4to el ajuste óptimo se obtiene con la función GEV.





Para añadir, si comparamos el desempeño del modelo exponencial entre los distintos puntos para la variable TWL (*tabla 7*), se observa que los errores de los ajustes respecto a los valores empíricos son considerablemente mayores en zonas con ciclones tropicales, sobre todo para la zona del Golfo de México, lo cual era un resultado esperable considerando la influencia que los ciclones tropicales tienen en estas zonas. Sin embargo, teniendo esto en cuenta, llama la atención que para la otra función que cumple que $\xi = 0$ (Gumbel), tanto para TWL como el η , es la función óptima en 5 de los 21 puntos. Zonas tanto con ciclones tropicales como extra-tropicales.

Cabe mencionar también que, para los casos de los puntos 6 (Este de Norteamérica) para el TWL, y 6 y 11 para el η , en el que la función GEV y Gumbel devuelven el mismo error para la variable , para el primer caso se consideró como óptima la función GEV debido a que según el análisis de la significancia se obtenía que ξ era significativamente distinto de cero, mientras que para los otros casos se consideró que el óptimo era la función Gumbel, debido a que no era significativo.

6. Conclusiones

Como se comentó en el apartado de resultados, las zonas donde se observa un mayor valor del parámetro de forma son en el Golfo de México, el cuál es conocido por la frecuente presencia anual de ciclones tropicales.

Los modelos que se basan en el método POT (GPD y Exponencial) generalmente presentan un mayor RRMSE que los que emplean el método de selección por máximos anuales para los valores más extremos de las muestras. En otras palabras, estos modelos son los que peor se ajustan en las partes altas de la colas.

En tercer lugar, los ajustes a partir de distribuciones mixtas son las que mejor modelan los valores extremos de NTR en todos los puntos, exceptuando en el Mar del Norte. Estas funciones presentan un elevado nivel de precisión en los resultados, como ya se ha visto en las gráficas de periodos de retorno mostradas y las que se añaden en el anexo, ajustando con considerablemente buenos resultados tanto las partes bajas como las más altas de la distribución de las muestras.

Las distribuciones mixtas presentan una mejora considerable dentro de los modelos que se basan en el métodos de selección por bloques de tiempo. Esto se cumple tanto para los puntos localizados en zonas extra-tropicales, como los que tienen influencia de ciclones tropicales, por lo que parecen adecuadas a la hora de modelar variables que presentan más de una familia de valores extremos.

Se puede mejorar la fiabilidad de los resultados de este estudio revisando y repitiendo algunos pasos de la metodología. Se propone revisar la selección del umbral aplicando el método descrito para ello de manera que se favorezca la fiabilidad de los ajustes con el método POT. También se puede revisar el número de valores extremos a partir de los cuáles se calcula el RRMSE, para aumentar la cantidad de valores que engloba, incrementando así la robustez de esta métrica.

Por último, para futuras aplicaciones de las funciones mixtas, se propone investigar con diferetes combinaciones de funciones de distribución para aplicarlas en zonas en las que no se obtuvo como resultado que este ajuste era el óptimo.













Referencias

- Allsop, W., Bruce, T., Pearson, J., Franco, L., Burgon, J. & Ecob, C. (2005). Safety under wave overtopping – How overtopping processes and hazards are viewed by the public. 29th International Conference on Coastal Engineering.
- Bernardara, P., Mazas, F., Kergadallan, X., and Hamm, L.: A two-step framework for over-threshold modelling of environmental extremes, Nat. Hazards Earth Syst. Sci., 14, 635–647, https://doi.org/10.5194/nhess-14-635-2014, 2014.
- Bouma, J.J., François, D., Schram, A. & Verbeke, T. (2009). Assessing socio-economic impacts of wave overtopping: An institutional perspective. *Coastal Engineering*, 56, 203–209. https://doi. org/10.1016/j.coastaleng.2008.03.008
- Caruso, M. F. & Marani, M. (2022). Extreme-coastal-water-level estimation and projection: a comparison of statistical methods. *Nat. Hazards Earth Syst. Sci.*, 22, 1109–1128, https://doi.org/10.5194/nhess-22-1109-2022.
- Chen, Q., Wang, L., Zhao, H. & Douglass, S.L. (2007). Prediction of storm surges and wind waves on coastal highways in hurricane-prone areas. *Journal of Coastal Research*, 23(5), 1304–1317. http://dx.doi.org/10.2112/05-0465.1
- Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics.
- Daniel Codiga (2023). UTide Unified Tidal Analysis and Prediction Functions (https://www.Mathworks .com/matlabcentral/fileexchange/46523-utide-unified-tidal-analysis-and-prediction-functions), *MATLAB Central File Exchange*. Recuperado Julio 24, 2023.
- Davison, A.C. & Smith, R.L. (2018). Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society: Series B*, 52, 3, 393–425, https://doi.org/10.1111/j.2517-6161.1990.tb01796.x
- Dean, R.G. & Walton, T.L. (2009). Wave setup, handbook of coastal and ocean engineering. World Scientific, 1-23. https://doi.org/10.1142/9789812819307_0001
- Di Paola, G., Rodríguez, G. & Rosskopf, C.M. (2020). Short- to mid-term shoreline changes along the southeastern coast of Gran Canaria Island (Spain). *Rendiconti Lincei. Scienze Fisiche e Naturali*, 31, 89–102. https://doi.org/10.1007/s12210-020-00872-3
- Galambos, J. (1994). Extreme Value Theory for Applications. Extreme Value Theory and Applications. *Springer*, Boston, MA. https://doi.org/10.1007/978-1-4613-3638-9_1
- Godin, G. (1972). The Analysis of Tides. Toronto University Press.
- Guerra-Medina, D. & Rodríguez, G. (2021). Spatiotemporal Variability of Extreme Wave Storms in a Beach Tourism Destination Area. *Geosciences*, 11(6), 237. https://doi.org/10.3390/geosciences11060237





- Hunter, J. (2010). Estimating sea-level extremes under conditions of uncertain sea-level rise. *Climatic Change*, 99, 331–350. DOI 10.1007/s10584-009-9671-6
- Idier, D., Bertin, X., Thompson, P. & Pickering, M.D. (2019). Interactions between mean sea level, tide, surge, waves and flooding: Mechanisms and contributions to sea level variations at the coast. Surveys in Geophysics, 40(6), 1603–1630. https://doi.org/10.1007/s10712-019-09549-5
- Katz, R.W. (2002). Stochastic hurricane modeling of damage. Journal of Applied Meteorology and Climatology, 41. 754-762. https://doi.org/10.1175/1520-0450(2002)041<0754:SMOHD>2.0.CO;2
- Knott, J.F., Elshaer, M., Daniel, J.S., Jacobs, J.M. & Kirshen, P. (2017). Assessing the effects of rising groundwater from sea level rise on the service life of pavements in coastal road infrastructure. *Transportation Research Record: Journal of the Transportation Research Board*, 2639, 1-10. https://doi.org/10.3141/2639-01.
- Macdonald, E. (2018). Urban Waterfront Promenades. Routledge. https://www.scribd.com/document/ 447181905/Macdonald-Elizabeth-Urban-Waterfront-Promenades-Routledge-2018
- Menéndez, M. (2008). *Metodología para el análisis estadístico no estacionario de valores extremos de variables geofísicas*. Universidad de Cantabria.
- Menéndez, M. & Losada, I.J. (2017). GOW2: A global wave hindcast for coastal applications. *Coastal Engineering*, 124, 1-11. https://doi.org/10.1016/j.coastaleng.2017.03.005
- Muis, S., Verlaan, M., Winsemius, H. (2016). A global reanalysis of storm surges and extreme sea levels. *Nat Commun*, 7, 11969. https://doi.org/10.1038/ncomms11969
- Parker, B.B. (2007) Tidal analysis and prediction. Silver Spring, MD, NOAA NOS Center for Operational Oceanographic Products and Services, 378pp (NOAA Special Publication NOS CO-OPS 3). DOI: http://dx.doi.org/10.25607/OBP-191
- Pickands, J. (1975) Statistical Inference Using Extreme Order Statistics. *Annals of Statistics*, 3, 119-131. http://dx.doi.org/10.1214/aos/1176343003
- Pinet, P.R. (2003). Invitation to Oceanography. Boston: Jones & Bartlett Learning.
- Pugh, D. & Woodworth, P. (2014). Sea-level science: Understanding tides, surges, tsunamis and mean sealevel changes. Cambridge University Press.
- Robinson, R. & Thagensen, B. (2004). Road engineering development. Spon Press, (2), 544.
- Rodríguez, G., Nistal, A., Pérez, B. & López, A.F. (1999). Joint occurrence of high tide, surge and stormwaves on the northwest Spanish coast. Boletín Instituto Español de. Oceanografía, 15(1-4), 21-29.
- Rosbjerg, D., Madsen, H. & Rasmussen, P.F. (1992). Prediction in partial duration series with generalized pareto-distributed exceedances. *Water Resources Research*, 28, 3001-3010. https://doi.org/10.1029/92WR01750





- Ruggiero, P., Komar, P.D., Allan, J.C. (2010). Increasing wave heights and extreme value projections: The wave climate of the U.S. Pacific Northwest. *Coastal Engineering*, 57, 5, 539-552 https://doi.org/10.1016/j.coastaleng.2009.12.005
- Semedo, A., K. Sušelj, A. Rutgersson, and A. Sterl, 2011: A Global View on the Wind Sea and Swell Climate and Variability from ERA-40. J. Climate, 24, 1461–1479, https://doi.org/10.1175/2010JCLI3718.1.
- Smith, G. & Juria, N. (2019). Diagnosis of historical inundation events in the Marshall Islands to assist early warning systems. *Natural Hazard*, 99(1), 189–216. https://doi.org/10.1007/s11069-019-03735-9
- Stockdon, H.F., Holman, R.A., Howd, P.A. & Sallenger Jr, A.H. (2006). Empirical parameterization of setup, swash, and runup. *Coastal Engineering*. 53 (2006) 573–588. https://doi.org/10.1016/ j.coastaleng.2005.12.005 Sundar, V. (2021). Ocean wave dynamics for coastal and marine structures. World Scientific, 52, 364. https://doi.org/10.1142/12268
- Svendsen, I.A (2006). *Introduction to nearshore hydrodynamics*. Advanced Series on Ocean Engineering. https://kupdf.net/download/introduction-to-nearshore-hydrodynamics59060dc6dc0d603744959e8 e_pdf
- Vega, C.C. (2018). Estadística de eventos extremos en sistemas complejos. Universidad de Cantabria.
- Vega, J.L., González, J., & Rodríguez, G. (2013). Statistical assessment of annual patterns in coastal extreme wave conditions. *WIT Transactions on Ecology and The Environment*, Vol 169, 39-49.
- Vignudelli, S., Scozzari, A., Abileah, R., Gómez-Enri, J., Benveniste, J. & Cipollini, P. (2019). Water surface elevation in coastal and inland waters using satellite radar altimetry. *Elsevier*, Chapter 4 (pp. 87-127), doi:10.1016/B978-0-12-814899-0.00004-3.
- Weisse, R. & von Storch, H. (2010). Marine Climate and Climate Change: Storms, Wind Waves and Storm Surges. Springer. https://doi.org/10.1007/978-3-540-68491-6
- Woodworth, P.L., Hunter, J.R. Marcos, M., Caldwell, P., Menendez, M. and Haigh, I. 2017. Towards a global higher-frequency sea level data set. Geoscience Data Journal, 3, 50-59, doi:10.1002/gdj3.42.





Anexo

	Contribución NTR (0-1)	Contribución Set-up (0-1)	Rango de marea (m)	Mes eventos máximos	Tendencia lineal	СТ
St Petersburgo, USA (Florida)	0.254	0.172	1.128	10	1	Sí
Pensacola, USA (Alabama)	0.164	0.307	0.937	9	1	Sí
Galveston, USA (Texas)	0.268	0.268	1.143	9	1	Sí
Magueyes, Puerto Rico	0.010	0.585	0.442	2	1	Sí
Montauk, USA (Nueva York)	0.236	0.452	1.235	12	1	Sí
Lewes, USA (Delaware)	0.262	0.313	1.969	10	1	Sí
Cape May, USA (Delaware)	0.181	0.184	2.17215	10	1	Sí
New London, USA (Connectica)	0.250	0.374	1.37	12	1	Sí
Cuxhaven, Alemania	0.249	0.032	4.13	1	0	No
Naha, Japón	0.090	0.440	2.13	8	1	Sí
Miyakejima, Japán	0.125	0.528	2.1	9	1	Sí
Wakkanai, Japón	0.120	0.542	0.66	11	1	Sí
Cananeia, Brasil	0.138	0.256	1.7564	8	1	No
Apra, Guam	0.030	0.685	1.099	10	1	Sí
Weipa, Australia	0.248	0.8517	2.579	1	1	Sí
Pago Ago, USA	0.012	0.601	1.234	5	1	Sí
Wellington, NZ	0.071	0.589	1.443	5	1	Sí
Polinesia francesa, Francia	0.053	0.649	0.42815	5	1	Sí
Pointe de Galets, Reunión	-0.073	0.85177	0.819	7	1	Sí
Yakutat, USA	0.072	0.471	3.917	1	0	No
San Francisco, USA	-0.011	0.467	2.175	1	1	No





Figura 17. Mapa con el mes del año con el mayor número de eventos extremos asociados por punto.





DMCE			TWL		
RMSE	GEV	Gumbel	Mixtas	GPD	Exponencial
St Petersburgo, USA (Florida)	0.050	0.065	0.067	0.113	0.254
Pensacola, USA (Alabama)	0.263	0.249	0.165	0.429	0.579
Galveston, USA (Texas)	0.113	0.236	0.070	0.338	0.557
Magueyes, Puerto Rico	0.053	0.128	0.030	0.141	0.335
Montauk, USA (Nueva York)	0.123	0.152	0.062	0.268	0.346
Lewes, USA (Delaware)	0.062	0.062	0.077	0.189	0.288
Cape May, USA (Delaware)	0.022	0.030	0.027	0.089	0.069
New London, USA (Connectica)	0.135	0.157	0.022	0.293	0.322
Cuxhaven, Alemania	0.066	0.070	0.050	0.127	0.178
Naha, Japón	0.110	0.126	0.042	0.174	0.213
Miyakejima, Japán	0.157	0.147	0.030	0.275	0.325
Wakkanai, Japón	0.042	0.038	0.024	0.095	0.051
Apra, Guam	0.020	0.014	0.023	0.059	0.047
Cananeia, Brasil	0.086	0.103	0.052	0.123	0.288
Weipa, Australia	0.025	0.018	0.019	0.068	0.098
Pago Ago, USA	0.022	0.023	0.005	0.023	0.128
Wellington, NZ	0.093	0.190	0.066	0.317	0.402
Polinesia francesa, Francia	0.048	0.035	0.050	0.137	0.156
Pointe de Galets, Reunión	0.035	0.018	0.036	0.084	0.043
Yakutat, USA	0.035	0.136	0.027	0.069	0.339
San Francisco, USA	0.055	0.047	0.071	0.142	0.151

Tabla 7. RRMSE obtenido del ajuste de los empíricos de TWL con cada modelo para las 21 localizaciones.





DMCE			η		
RMSE	GEV	Gumbel	Mixtas	GPD	Exponencial
St Petersburgo, USA (Florida)	0.021	0.041	0.018	0.032	0.079
Pensacola, USA (Alabama)	0.451	0.229	0.232	0.215	0.268
Galveston, USA (Texas)	0.185	0.123	0.283	0.389	0.314
Magueyes, Puerto Rico	0.102	0.140	0.042	0.160	0.265
Montauk, USA (Nueva York)	0.115	0.162	0.067	0.187	0.288
Lewes, USA (Delaware)	0.046	0.046	0.063	0.064	0.064
Cape May, USA (Delaware)	0.049	0.046	0.015	0.086	0.067
New London, USA (Connectica)	0.079	0.115	0.040	0.109	0.193
Cuxhaven, Alemania	0.060	0.094	0.039	0.082	0.131
Naha, Japón	0.093	0.100	0.076	0.093	0.167
Miyakejima, Japán	0.148	0.148	0.153	0.229	0.233
Wakkanai, Japón	0.047	0.029	0.029	0.072	0.086
Apra, Guam	0.014	0.015	0.016	0.026	0.060
Cananeia, Brasil	0.096	0.122	0.114	0.144	0.219
Weipa, Australia	0.018	0.035	0.025	0.029	0.111
Pago Ago, USA	0.020	0.023	0.025	0.037	0.039
Wellington, NZ	0.149	0.242	0.078	0.338	0.467
Polinesia francesa, Francia	0.051	0.034	0.054	0.134	0.103
Pointe de Galets, Reunión	0.057	0.047	0.043	0.127	0.112
Yakutat, USA	0.054	0.051	0.047	0.121	0.087
San Francisco, USA	0.069	0.053	0.074	0.138	0.146

Tabla 8. RRMSE obtenido del ajuste de los empíricos de η con cada modelo para las 21 localizaciones.





Modelo óptimo	NTR	TWL	η
St Petersburgo, USA (Florida)	Mixtas	GEV	Mixtas
Pensacola, USA (Alabama)	Mixtas	Mixtas	GPD
Galveston, USA (Texas)	Mixtas	Mixtas	Gumbel
Magueyes, Puerto Rico	Mixtas	Mixtas	Mixtas
Montauk, USA (Nueva York)	Mixtas	Mixtas	Mixtas
Lewes, USA (Delaware)	Mixtas	GEV	Gumbel
Cape May, USA (Delaware)	Mixtas	GEV	Mixtas
New London, USA (Connectica)	Mixtas	Mixtas	Mixtas
Cuxhaven, Alemania	GEV	Mixtas	Mixtas
Naha, Japón	Mixtas	Mixtas	Mixtas
Miyakejima, Japán	Mixtas	Mixtas	GEV
Wakkanai, Japón	Mixtas	Mixtas	Mixtas
Apra, Guam	Mixtas	Gumbel	GEV
Cananeia, Brasil	Mixtas	Mixtas	GEV
Weipa, Australia	Mixtas	Gumbel	GEV
Pago Ago, USA	Mixtas	Mixtas	GEV
Wellington, NZ	Mixtas	Mixtas	Mixtas
Polinesia francesa, Francia	Mixtas	Gumbel	Gumbel
Pointe de Galets, Reunión	Mixtas	Gumbel	Mixtas
Yakutat, USA	Mixtas	Mixtas	Mixtas
San Francisco, USA	Mixtas	Gumbel	Gumbel

Tabla 8. Modelos óptimos resultantes del análisis para cada uno de los 21 puntos.