

Disipación y Energía de los Ciclones
Tropicales: Ajustes y Test de Bondad de
Ajuste

Oliver Planes
TUTOR: Álvaro Corral

Agradecimientos

A mis compañeros y profesores por su humanidad y afecto.

A mi tutor, Álvaro Corral, del que he recibido apoyo, orientación y consejo durante todo el proyecto.

Índice general

I Disipación y Energía de los Ciclones Tropicales: Ajustes y Test de Bondad de Ajuste	4
1. Introducción	5
2. Material y métodos	10
2.1. Datos	10
2.2. Compleción de las bases de datos	12
2.3. Distribución de probabilidad del PDI, IEI y OEI	16
2.4. Estimación de la ley de potencias	20
2.5. Determinación del intervalo de ajuste	22
2.6. Contraste de Bondad de Ajuste de Kolmogorov-Smirnov . . .	22
2.7. Cálculo del p-valor de KS	23
2.8. Comparación de las distribuciones de probabilidad	24
3. Resultados	26

4. Conclusiones	32
II Anexo	35
1. Relación entre el OEI, el IEI y la energía de un ciclón tropical	36
2. Código fuente en R	38

Parte I

Disipación y Energía de los Ciclones Tropicales: Ajustes y Test de Bondad de Ajuste

1 Introducción

¿Qué son los ciclones?

Los *ciclones* son sistemas tormentosos que circulan alrededor de un centro de baja presión. La palabra ciclón proviene del griego *κυκλων* (kyklon) que significa círculo en movimiento, haciendo referencia a fenómenos meteorológicos caracterizados por ser un sistema de tormentas en circulación cerrada alrededor de un centro de baja presión. Pueden formarse en varias regiones del planeta, distinguiendo entre los *cyclones tropicales*, *cyclones extratropicales*, *cyclones subtropicales* y *cyclones polares*. El presente trabajo se centra en los ciclones tropicales, que son aquellos que se forman en las regiones tropicales del planeta y que se caracterizan por ser altamente devastadores.

Clasificación de los ciclones tropicales

Según la velocidad del viento los ciclones tropicales pueden denominarse *depresión tropical*, *tormenta tropical* y *huracán o tifón*.

En el caso de las depresiones tropicales, la velocidad sostenida máxima del viento (*maximum sustained wind speed*) es inferior a los 34 nudos¹. Para las tormentas tropicales la velocidad sostenida máxima es supera los 33 nudos y no alcanza los 63. Finalmente, los huracanes y los tifones poseen una velocidad sostenida máxima del viento superior a los 62 nudos.

Esencialmente, los huracanes y los tifones son el mismo fenómeno, simplemente cambia su denominación según la región geográfica donde se forman. Las zonas que se tratarán en este trabajo son el norte del Océano Atlántico

¹1 nudo = $1852 \text{ m} \text{h}^{-1} \approx 0.5144 \text{ m} \text{s}^{-1}$

Categoría	v_{\max}
1	64 – 82 nudos $119 - 153 \text{ kmh}^{-1}$
2	83 – 95 nudos $154 - 177 \text{ kmh}^{-1}$
3	96 – 112 nudos $178 - 208 \text{ kmh}^{-1}$
4	113 – 136 nudos $209 - 251 \text{ kmh}^{-1}$
5	≥ 137 nudos $\geq 252 \text{ kmh}^{-1}$

Cuadro 1.1: Escala de Saffir-Simpson actualizada a febrero de 2012. La medición de la velocidad máxima del viento es sostenida durante un minuto, es decir, se promedia la velocidad durante 60 segundos.

y el noreste del Océano Pacífico, que son zonas de huracanes. Dentro de la categoría de huracanes, éstos pueden clasificarse[20] a su vez de acuerdo a la Escala de Saffir-Simpson (ver cuadro 1.1).

Influencia de los huracanes en la vida humana. Importancia de su estudio.

Los huracanes son los fenómenos climatológicos terrestres más devastadores, razón por la que su estudio resulta de suma importancia. No es fácil estudiar los fenómenos ciclónicos ya que no podemos ponerlos en un laboratorio. Para poder estudiarlos son fundamentales las mediciones periódicas de sus características debido a que tales mediciones posibilitan un tratamiento estadístico.

La influencia de los ciclones tropicales sobre la vida humana es tal, que incluso se han realizado estudios para valorar el impacto que tienen en la economía, por ejemplo en la tasa de crecimiento económico local [23]. De todos modos, cuando se toma conciencia de tal influencia es cuando se ven los daños causados directamente sobre la población, sus bienes y entorno. A modo de ejemplos, en diciembre de 1974 el huracán *Tracy* arrasó la ciudad de Darwin, en la costa septentrional australiana, dejando tras de sí 71 muertos



ISS024E012814

Figura 1.1: Imagen del Huracán Danielle captada el 27 de agosto de 2010 desde la Estación Espacial Internacional. **Fuente:** Douglas H. Wheelock, Coronel del ejército de los EEUU. Formó parte de la tripulación de la misión Soyuz TMA-19 en la Estación Espacial Internacional entre el 15 de junio y el 25 de noviembre de 2010. http://twitter.com/Astro_Wheels

y 650 heridos. El 52 % de las casas fue destruido y el 80 % de los edificios residenciales fueron totalmente destrozados o quedaron inhabitables [6]. Más recientemente, en agosto de 2005, el huracán *Katrina* asoló el golfo de México dejando tras de sí más de 1800 fallecidos, 250000 desplazados y unos daños valorados en 125000 millones de dólares [16]. El *Tracy* y el *Katrina* son ejemplos moderados en cuanto a los efectos de los ciclones. Registros históricos recogen estimaciones de hasta cientos de miles de personas fallecidas por los efectos de huracanes:

Al huracán de Kyushu ² (Japón) en el año 1281 se le atribuyen más de 100000 víctimas entre las que se encontraba gran parte de la flota mongola, que estaba en plena invasión de Japón y perdió unos 13000 integrantes.

El huracán de Backerganj de 1976 (Bangladesh en la actualidad) dejó 200000 víctimas mortales, tal y como hiciera otro huracán en la misma zona casi 300 años antes. Aparte de las victimas directas por el huracán, se estima que otras 100000 personas murieron por los efectos del hambre consecuencia de la catástrofe.

El fenómeno ciclónico al que mayor número de víctimas se atribuye tuvo lugar en 1970 y se llamó huracán *Bhola*. Afectó a Bangladesh y al este de la India causando aproximadamente medio millón de muertes. A pesar de

²Se le llamó *Kamikaze* que significa viento divino y supone el origen del mito según el cual Japón era el país elegido por los dioses y por tanto estaba bajo su protección.

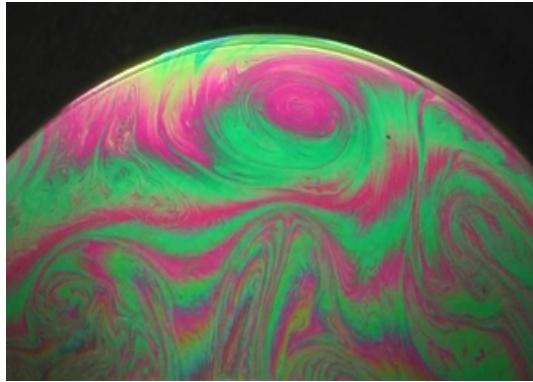


Figura 1.2: No es factible introducir un huracán en un laboratorio, pero en una semiesfera definida por una burbuja de jabón se pueden generar vórtices aislados al calentar la zona del ecuador. Estos vórtices poseen algunas características similares a las de los ciclones terrestres [21]. **Fuente:** Centre national de la recherche scientifique. <http://www2.cnrs.fr/presse/communique/1322.htm>

la devastación que causó, el *Bhola* 'solamente' alcanzó la categoría 3 de la Escala de Saffir-Simpson de las 5 posibles (ver cuadro 1.1).

Con la intención de mitigar el efecto destructivo de los ciclones tropicales, el gobierno de los Estados Unidos de América puso en funcionamiento en el año 1962 el proyecto *Stormfury* [24]. El proyecto consistía en inyectar yoduro de plata en el núcleo del huracán de modo que éste aumentara su radio y se debilitara. En un inicio parecía que el método era eficaz, pero tras la disminución momentánea de la velocidad del viento, el huracán volvía a ganar intensidad. Además las conclusiones del proyecto indicaban que no era seguro que las modificaciones de la intensidad de los huracanes se debieran fundamentalmente a la inyección del yoduro de plata.

Según parece, la manipulación del clima no permite paliar significativamente los efectos de los huracanes en la actualidad. El énfasis en el estudio de los fenómenos ciclónicos puede ayudar a mejorar esta manipulación así como nuestra capacidad predictiva.

Calentamiento global y huracanes. ¿Cómo se relacionan?

La teoría de formación de ciclones tropicales induce a pensar que un aumento en la temperatura media global implica un aumento en la intensidad de éstos (estableciendo una relación directa entre la intensidad de tales fenómenos y el calentamiento global), pero los estudios basados en la frecuencia de aparición de huracanes no mostraban ninguna tendencia. Por esta razón K. Emanuel definió en 2005 el *Power Dissipation Index* (PDI), como medida de destructividad potencial de un huracán. Mediante la tendencia de aumento del PDI a partir de 1970 Emanuel argumentaba una tendencia de aumento en la intensidad de los huracanes [14].

La pregunta es, ¿unos PDI crecientes a lo largo de esos años implican realmente una tendencia de aumento?

La respuesta a esta pregunta nos indica si el calentamiento global afecta a la génesis y destructividad de los huracanes tropicales. Para intentar encontrar una respuesta, el objetivo del presente proyecto es la modelización de la distribución probabilística del PDI en las zonas del norte del Océano Atlántico y del noreste del Océano Pacífico así como de otros dos índices que serán definidos más adelante.

La intención es poder establecer distribuciones probabilísticas de los valores correspondientes a índices que midan la destructividad potencial de los ciclones tropicales a partir de los datos registrados de forma independiente a cuándo han tenido lugar. Así pues este trabajo es un primer paso para posteriormente estudiar la influencia del calentamiento global.

En caso de poder modelizar las distribuciones de probabilidad mencionadas, se procederá a estudiar las relaciones que se puedan establecer entre los índices tratados.

2 Material y métodos

2.1. Datos

Los datos utilizados en el presente proyecto corresponden a la base de datos sobre ciclones tropicales del Océano Atlántico y del Nordeste del océano Pacífico del *Regional and Mesoscale Meteorology Branch* (RAMMB)[1] que pertenece al *National Oceanic and Atmospheric Administration* (NOAA)[2]; más concretamente corresponden al *National Hurricane Center* (NHC)[4], y pueden ser consultados en la página web:

rammb.cira.colostate.edu/research/tropical_cyclones/tc_extended_best_track_dataset.

Los Documents que contienen toda la información poseen mediciones sobre posición (ver figura 2.1), velocidad y tamaño tomadas cada seis horas, concretamente a las 00.00, 06.00, 12.00 y 18.00 horas UTC (*Universal Time Coordinated*).

Los registros sobre ciclones tropicales del Atlántico datan de 1988 al 2010, mientras que los del Pacífico comprenden el periodo de 2001 al 2010.

La base de datos referente al norte del Océano Atlántico consta de 9691 registros correspondientes a un total de 303 ciclones tropicales. La base de datos con los registros del noreste del Océano Pacífico tiene 3791 entradas pertenecientes a 140 ciclones.

De todos los campos que conforman las bases disponibles, son de especial importancia la máxima velocidad del viento, el radio de máxima velocidad del viento y el radio de la isobara cerrada exterior. La información de las máximas velocidades es completa, pero respecto a los radios falta una cantidad relevante de información. Para paliar el efecto de la falta de datos se procederá a completarlos mediante técnicas de interpolación.

Para el tratamiento de los datos así como para la programación de las rutinas

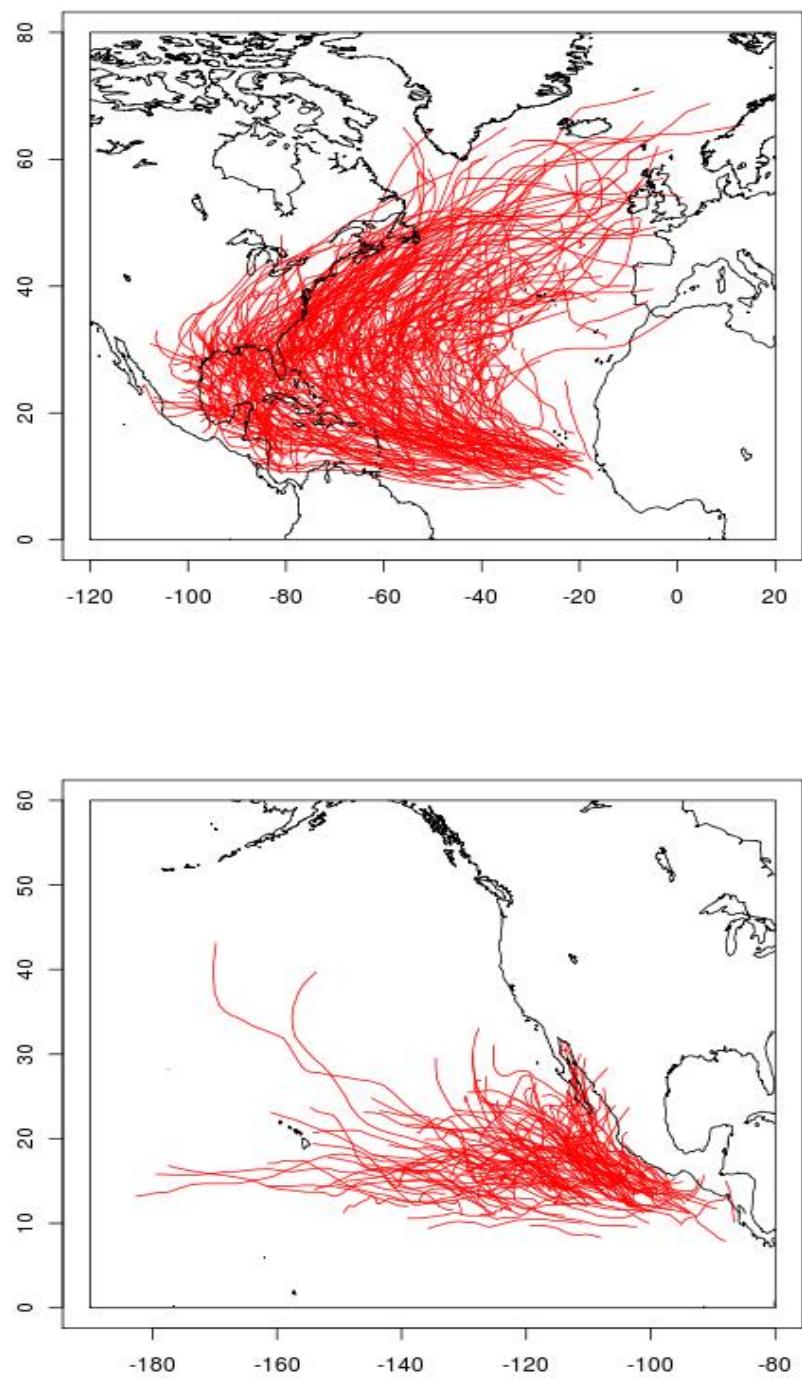


Figura 2.1: Trayectorias seguidas por los distintos ciclones tropicales que conforman las bases de datos.

informáticas desarrolladas se ha utilizado el entorno R [3]. (Ver anexo para consultar el código fuente)

2.2. Compleción de las bases de datos

Dado un ciclón concreto, se puede considerar que su radio de máxima velocidad del viento (todo el desarrollo es análogo para el radio de la isobara cerrada exterior) es una función que depende del tiempo:

$$\begin{aligned} r : [T_0, T_1] &\rightarrow \mathbb{R}_+ \\ t &\mapsto r(t) \end{aligned}$$

La situación real es que la base de datos en la que se encuentra el ciclón contiene puntos de esta función, así pues si se supone que $r(t)$ es continua, parece natural interpolar tales puntos para obtener una expresión polinómica $\hat{r}(t)$ que aproxime $r(t)$ y que permita valorar r en abscisas t no registradas en la base de datos.

Si se pretende interpolar la función $r(t)$ en los puntos disponibles mediante una función polinómica, la naturaleza oscilatoria de los polinomios de grados altos y la característica de que una fluctuación en una porción pequeña del intervalo de definición puede inducir fluctuaciones muy grandes en todo el intervalo, parecen desaconsejar este método. Para salvar esta problemática se puede recurrir a la interpolación polinómica fragmentaria, que consiste en considerar una partición del intervalo de definición a partir del conjunto de abscisas de los puntos de interpolación y aplicar interpolación polinómica utilizando polinomios de grados bajos a cada uno de estos intervalos.

El caso más simple de interpolación polinómica fragmentaria consiste en considerar la partición obtenida mediante todas las abscisas disponibles y construir el polígono que une sus imágenes, es decir, interpolar puntos consecutivos mediante segmentos. El problema de considerar esta interpolación es que no se garantiza la diferenciabilidad en cada uno de los extremos de los subintervalos, propiedad que sería deseable en este contexto ya que las funciones que modelizan fenómenos físicos suelen gozar de ella.

Si se intenta unir los nodos mediante polinomios de segundo grado, tampoco se puede asegurar tal diferenciabilidad debido a qué tales polinomios vienen determinados por tres coeficientes. Al imponer las condiciones de interpolación y diferenciabilidad en uno de estos polinomios, se construye un sistema lineal sobre-determinado (4 condiciones para determinar 3 coeficientes), dando lugar a posibles sistemas incompatibles (o lo que es lo mismo, que no

existen tales polinomios).

La interpolación fragmentaria más común y que garantiza de forma certera la diferenciabilidad de la función $\hat{r}(t)$ en todo su dominio de definición es la interpolación de trazador cúbico, consistente en considerar la misma partición que en el caso anterior pero uniendo puntos consecutivos mediante polinomios cúbicos. Aparte del grado de los polinomios se impone que la derivada y la segunda derivada de los trazadores cúbicos coincidan en extremos comunes de los subintervalos donde se definen. Imponer igualdad de las segundas derivadas en los nodos además de condiciones frontera permiten obtener un sistema lineal compatible y determinado (y en consecuencia la interpolación es única):

Consideremos t_1, \dots, t_n abscisas de interpolación y sea $S_j(t)$ el polinomio cúbico interpolador en $[t_j, t_{j+1}]$ para cada $j = 1, \dots, n - 1$. Se ha de cumplir

1. $S_j(t_j) = r(t_j)$ para cada $j = 1, \dots, n - 1$.
2. $S_{j+1}(t_{j+1}) = S_j(t_{j+1})$ para cada $j = 1, \dots, n - 2$.
3. $S'_{j+1}(t_{j+1}) = S'_j(t_{j+1})$ para cada $j = 1, \dots, n - 2$.
4. $S''_{j+1}(t_{j+1}) = S''_j(t_{j+1})$ para cada $j = 1, \dots, n - 2$.

Si además se satisface la siguiente condición de frontera

5. $S''_1(t_1) = S''_{n-1}(t_n) = 0$

la función de interpolación se llama *adaptador natural* [8]

Una vez determinados los trazadores cúbicos $S_i(t)$ con $i = 1, \dots, n - 1$, se define la función de interpolación o adaptador cúbico natural mediante:

$$S(t) = \begin{cases} S_1(t) & \text{si } t \in [t_1, t_2) \\ S_2(t) & \text{si } t \in [t_2, t_3) \\ \vdots & \\ S_i(t) & \text{si } t \in [t_i, t_{i+1}) \\ \vdots & \\ S_{n-2}(t) & \text{si } t \in [t_{n-2}, t_{n-1}) \\ S_{n-1}(t) & \text{si } t \in [t_{n-1}, t_n] \end{cases} \quad (2.1)$$

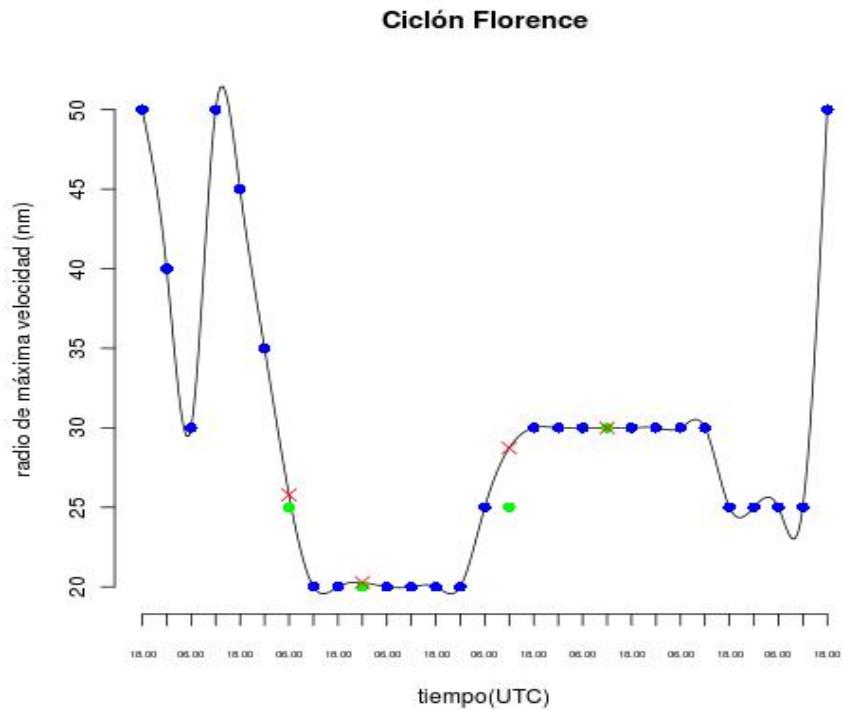


Figura 2.2: Interpolación mediante trazadores cúbicos de las mediciones del radio de máxima velocidad del ciclón Florence tomadas entre las 18.00 horas del 10 de Septiembre del 2000 a las 18.00 horas del 19 de Septiembre del mismo año en el norte del Océano Atlántico. En el caso del ciclón Florence, la base de datos es completa pero, a modo de ejemplo, se han extraído 4 mediciones quedando el registro completo en un 82 % y se ha procedido a completarlo. La función representada es la función de interpolación. Los puntos azules y verdes corresponden a las medidas correctas de la base de datos. Las equis rojas indican los valores interpolados que sustituyen a los valores exactos marcados en verde.

Determinada la función de interpolación $S(t)$ en $I = [t_1, t_n]$, se define $\hat{r}|_I(t) = S(t)$.

Este proceso permite asignar valores bastante coherentes a cualquier abscisa t dentro del intervalo I para la cual no se disponga de su radio en la base de datos, pero puede ser que haya que completar valores fuera de ese intervalo, es decir, extrapolar.

En caso de querer asignar valores a instantes t fuera del intervalo I , consideraré las prolongaciones de $S_1(t)$ y de $S_{n-1}(t)$ por la izquierda y por la derecha respectivamente, por lo que defino

$$\bar{S}(t) = \begin{cases} S_1(t) & \text{si } T_0 \leq t < t_1 \\ S(t) & \text{si } t_1 \leq t \leq t_n \\ S_{n-1}(t) & \text{si } t_n < t \leq T_1 \end{cases} \quad (2.2)$$

Finalmente, imponiendo que se respete el recorrido de la función $r(t)$, puedo definir la función $\hat{r}(t)$ a utilizar para llenar los datos faltantes:

$$\hat{r}(t) = \max\{4, \bar{S}(t)\} \quad (2.3)$$

La razón de asignar a los radios un valor mínimo de 4 millas náuticas se debe a que es el valor mínimo registrado para un radio. Otra opción sería asignar un valor mínimo de 0 millas náuticas pero conceptualmente sería sinónimo de la desaparición del ciclón.

Una vez determinado el método de compleción de la base de datos, hay que tener en cuenta la fiabilidad de tal compleción. Claro está que si un ciclón posee una importante cantidad de registros, pongamos 20, y solamente falta un dato (registro completo en un 95 %), se completa el registro y se acepta sin reparos. En el otro extremo, si de los 20 datos se dispone sólo de 2 de ellos (registro completo en un 10 %), parece natural rechazar este ciclón a la hora de calcular índices en los que se utilice la información faltante ya que la compleción no será nada fiable al disponer de muy poca información a interpolar. Para casos intermedios la decisión puede ser más complicada y discutible, razón por la que la forma de proceder será imponer un límite de aceptación. En este caso el límite de aceptación impuesto a un ciclón para completar su registro será estar completo en un 80 %.

Un aspecto importante a tener en cuenta al utilizar este método es que puede generar un sesgo. (ver figura 2.6).

2.3. Distribución de probabilidad del PDI, IEI y OEI

El PDI de un ciclón tropical que tiene lugar entre los instantes t_0 y t_1 se define [14] como

$$PDI = \int_{t_0}^{t_1} v_{\max}^3 dt \quad (2.4)$$

pero a efectos de computación, al tener registros discretos cada seis horas, su cálculo se realiza mediante

$$PDI \approx \sum_{t_0}^{t_1} v_{\max}^3 \Delta t \quad (2.5)$$

Donde v_{\max} es la máxima velocidad del viento (en ms^{-1}) en cada instante de tiempo.

Por otro lado defino el IEI (*Inner Energy Index*) entre t_0 y t_1 como

$$IEI = \int_{t_0}^{t_1} v_{\max}^3 R_{\max}^2 dt \quad (2.6)$$

pero, del mismo modo que en el caso del PDI, su cálculo se realiza mediante

$$IEI \approx \sum_{t_0}^{t_1} v_{\max}^3 R_{\max}^2 \Delta t \quad (2.7)$$

siendo R_{\max} el radio (m) de máxima velocidad del viento (ver anexo).

Asimismo defino el OEI (*Outer Energy Index*) entre los instantes t_0 y t_1 mediante

$$OEI = \int_{t_0}^{t_1} v_{\max}^3 R_{\text{iso}}^2 dt \quad (2.8)$$

que como anteriormente su cálculo se realiza con la expresión

$$OEI \approx \sum_{t_0}^{t_1} v_{\max}^3 R_{\text{iso}}^2 \Delta t \quad (2.9)$$

con R_{iso} el radio (m) de la isóbara exterior cerrada.

La diferencia en su denominación radica en la magnitud de los radios, siendo los de máxima velocidad del viento menores que los de la isóbara exterior.

Sea X la lista de los PDI correspondientes a una de las bases de datos. El valor del PDI (también el valor del IEI y el del OEI) puede considerarse una variable aleatoria, así que sea $D(x)$ su función de densidad (respectivamente $F(x)$ su función de distribución). La densidad se puede definir en un valor y considerando un intervalo $[y, y + dy]$ con $dy \rightarrow 0^+$ mediante el cociente de la probabilidad de que el PDI esté en el intervalo dividida por la longitud del intervalo, es decir,

$$D(y) \equiv \frac{P(y \leq X < y + dy)}{dy} \quad (2.10)$$

La probabilidad del numerador se puede estimar mediante el número de valores contenidos en el intervalo, $n(y) = \#\{x \in X \mid y \leq x < y + dy\}$, dividido por el número total de valores, N , así pues

$$D(y)dy = P(y \leq X < y + dy) \approx \frac{n(y)}{N} \quad (2.11)$$

En la práctica es necesario considerar $dy > 0$. Para representar gráficamente la función densidad se utiliza el método conocido como *logarithmic binning*[17][18], consistente en considerar una partición en la que cada valor se obtiene multiplicando el anterior por una constante c obteniendo intervalos

$$I_1 = [m, cm), I_2 = [cm, c^2m), I_3 = [c^2m, c^3m), \dots$$

siendo m un valor inicial.

dy_i es la longitud de I_i . Obsérvese que $dy_i = c^{i+1}m - c^im = c^i(c - 1)m$. En este caso la constante escogida es $c = 10^{1/5}$ y valor inicial $m = 10^8$. Una vez construida la partición, como si de un histograma se tratase, se consideran los vértices del polígono de frecuencias (ver figuras 2.32.42.5).

La representación de las estimaciones correspondientes a las funciones de densidad indican que en una parte de su dominio éstas pueden ajustar a una recta, característica propia de las funciones potenciales, es decir, funciones de la forma $f(x) = Cx^{-\alpha}$, $C \in \mathbb{R}$ ya que $\log(f(x)) = \log(C) - \alpha \log(x)$.

Las desviaciones a la ley de potencias para valores bajos y altos de la variable pueden deberse, respectivamente, a datos incompletos y al tamaño finito de las cuencas.

Cabe destacar que las funciones de densidad de tipo potencial son muy utilizadas en fenómenos críticos autoorganizados [7], así como fenómenos físicos, biológicos, económicos, demográficos [18],...

Es inmediato ver que si una función densidad es de la forma $f(x) = Cx^{-\alpha}$ y está definida en un intervalo $[A, B]$, su distribución es:

$$F(x) = \int_A^x Cy^{-\alpha} dy = \frac{C}{1-\alpha} (x^{1-\alpha} - A^{1-\alpha})$$

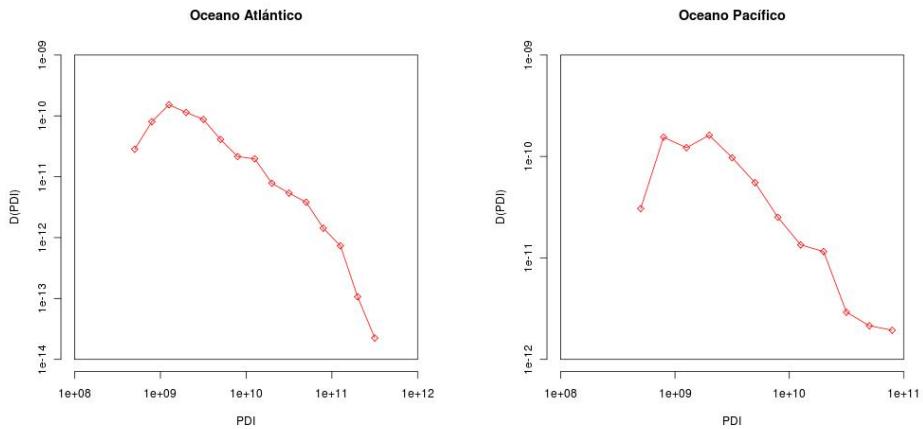


Figura 2.3: Estimación de la función densidad, $D(x)$, del PDI para el norte del Océano Atlántico y para el noreste del Pacífico.

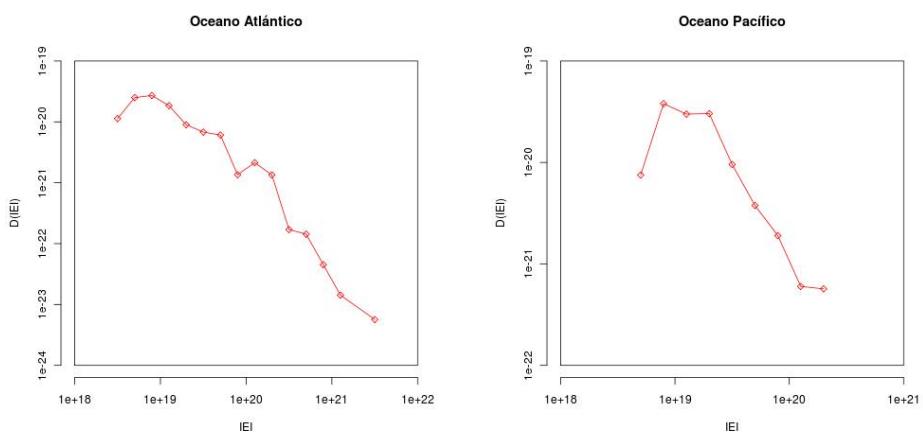


Figura 2.4: Estimación de la función densidad del IEI para el norte del Océano Atlántico y para el noreste del Pacífico.

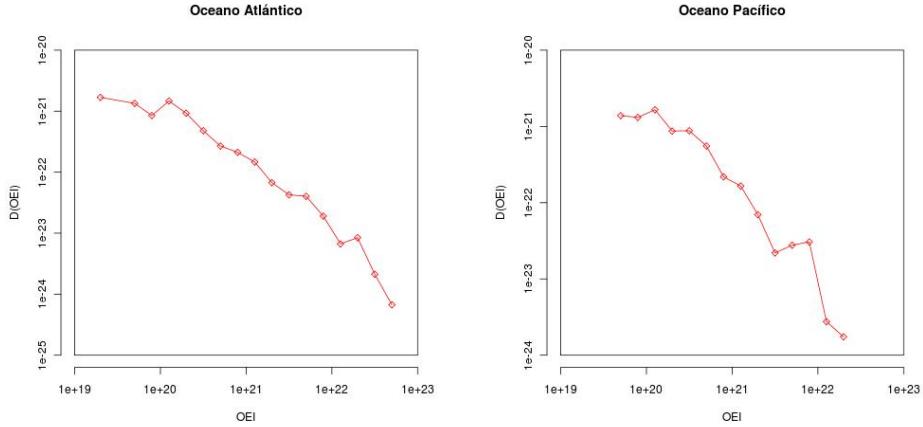


Figura 2.5: Estimación de la función densidad del OEI para el norte del Océano Atlántico y para el noreste del Pacífico.

y la función complementaria de la distribución es:

$$G(x) = \int_x^B Cy^{-\alpha} dy = \frac{C}{1-\alpha} (B^{1-\alpha} - x^{1-\alpha})$$

La constante C tiene efecto normalizador de la función $D(x)$, por lo que depende del intervalo de ajuste. Para determinar su valor simplemente hay que imponer que la ley de potencias $D(x)$ es una función de densidad definida en el intervalo $[A, B]$ y por tanto su integral en ese intervalo vale 1:

$$1 = \int_A^B Cx^{-\alpha} dx = \left[\frac{Cx^{1-\alpha}}{1-\alpha} \right]_A^B = C \frac{B^{1-\alpha} - A^{1-\alpha}}{1-\alpha} \Rightarrow C = \frac{1-\alpha}{B^{1-\alpha} - A^{1-\alpha}}$$

sea $r = \frac{A}{B}$. Se obtiene

$$C = \frac{1-\alpha}{\frac{1}{B^{\alpha-1}} - \frac{1}{A^{\alpha-1}}} = \frac{(1-\alpha)A^{\alpha-1}}{\left(\frac{A}{B}\right)^{\alpha-1} - 1} = \frac{(\alpha-1)A^{\alpha-1}}{1 - r^{\alpha-1}}$$

Se concluye que la función densidad es

$$D(x) = \frac{(\alpha-1)A^{\alpha-1}}{1 - r^{\alpha-1}} x^{-\alpha} \quad (2.12)$$

y por lo tanto

$$F(x) = \frac{x^{1-\alpha} - A^{1-\alpha}}{B^{1-\alpha} - A^{1-\alpha}} = \frac{1 - \left(\frac{A}{x}\right)^{\alpha-1}}{1 - r^{\alpha-1}} \quad (2.13)$$

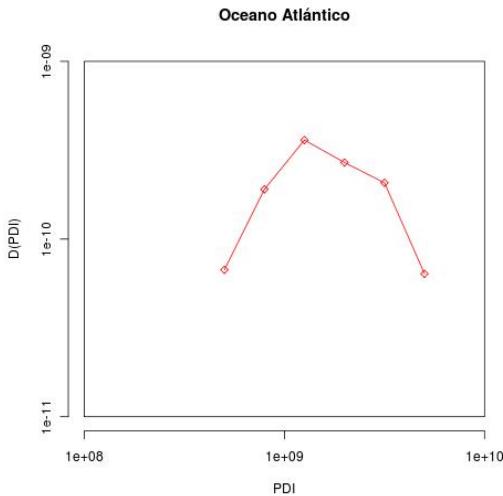


Figura 2.6: Aproximación de la función de densidad del PDI para el norte del Océano Atlántico considerando solamente aquellos ciclones que se utilizan para el cálculo del OEI (en la mayoría de ellos se ha completado el registro de datos). Se puede observar que se reproduce la parte correspondiente a los fenómenos con menor PDI (ver figura 2.3). La conclusión es que el método de compleción y selección de casos para la estimación de la distribución del OEI es sesgado. Esta conclusión es igualmente válida para el IEI ya que en la estimación de su distribución participan prácticamente los mismos ciclones.

$$G(x) = \frac{B^{1-\alpha} - x^{1-\alpha}}{B^{1-\alpha} - A^{1-\alpha}} = \frac{\left(\frac{A}{x}\right)^{\alpha-1} - r^{\alpha-1}}{1 - r^{\alpha-1}} \quad (2.14)$$

Tal y como pasa con la función de densidad, la función de distribución y su complementaria han de estimarse a partir de los datos disponibles mediante la función de distribución empírica (ver fórmula 2.20).

2.4. Estimación de la ley de potencias

Llegados a este punto, para determinar la ley de potencias hay que estimar el exponente α . La estimación de α depende de los datos comprendidos dentro del intervalo de ajuste, así que es natural buscar aquel exponente que sea el más plausible según los datos, es decir, α se estima por máxima verosimilitud.

Téngase en cuenta que el procedimiento utilizado para estimar gráficamente

las distribuciones no afecta a la estimación de máxima verosimilitud, simplemente se ha utilizado como ilustración visual.

Sea N_{AB} el número de datos comprendidos en el intervalo de ajuste $[A, B]$. La función de máxima verosimilitud es

$$V(x_1, \dots, x_{N_{AB}}; \alpha) = \prod_{i=1}^{N_{AB}} D(x_i) = C^{N_{AB}} \prod_{i=1}^{N_{AB}} x_i^{-\alpha}$$

y su logaritmo es

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_{N_{AB}}; \alpha) &= N_{AB} \log(C) - \alpha \sum_{i=1}^{N_{AB}} \log(x_i) \propto \log(C) - \frac{\alpha}{N_{AB}} \sum_{i=1}^{N_{AB}} \log(x_i) = \\ &= \log\left(\frac{(\alpha-1)A^{\alpha-1}}{1-r^{\alpha-1}}\right) - \frac{\alpha}{N_{AB}} \sum_{i=1}^{N_{AB}} \log(x_i) = \\ &= \log\left(\frac{\alpha-1}{1-r^{\alpha-1}}\right) + (\alpha-1) \log(A) - \frac{\alpha}{N_{AB}} \sum_{i=1}^{N_{AB}} \log(x_i) \end{aligned}$$

Fijados A i B se estima α mediante α_d cumpliendo

$$\forall \alpha, \mathcal{L}(x_1, \dots, x_{N_{AB}}; \alpha) \leq \mathcal{L}(x_1, \dots, x_{N_{AB}}; \alpha_d)$$

La varianza del estimador [5] viene dada por la fórmula

$$\sigma_{\alpha_d}^2 = \frac{1}{N_{AB}} \left\{ \frac{1}{(\alpha_d-1)^2} - \frac{r^{\alpha_d-1} \log(r)^2}{(1-r^{\alpha_d-1})^2} \right\}^{-1} \quad (2.15)$$

En el caso en que $B \rightarrow +\infty$ (y equivalentemente $r \rightarrow 0$) si $\alpha > 1$, la función de verosimilitud que se obtiene es

$$\lim_{r \rightarrow 0} \mathcal{L}(x_1, \dots, x_{N_{AB}}; \alpha) = \log(\alpha-1) + (\alpha-1) \log(A) - \frac{\alpha}{N_{\geq A}} \sum_{i=1}^{N_{\geq A}} \log(x_i) \quad (2.16)$$

cuyo máximo viene dado por

$$\alpha_d = 1 + \left\{ \left(\frac{1}{N_{\geq A}} \sum_{i=1}^{N_{\geq A}} \log(x_i) \right) - \log(A) \right\}^{-1}$$

y la varianza queda

$$\lim_{r \rightarrow 0} \sigma_{\alpha_d}^2 = \frac{(\alpha_d-1)^2}{N_{\geq A}} \quad (2.17)$$

En este caso la ley potencial no está truncada superiormente, aunque este caso particular no se utilizará para el ajuste. El no truncamiento impide encontrar el estimador analíticamente y el método deviene más complicado.

2.5. Determinación del intervalo de ajuste

En la sección previa se ha considerado fijo el intervalo de definición $[A, B]$, pero tales valores A y B han de ser determinados de alguna forma en el sentido de que la ley de potencias se ajuste “al máximo” en él. Es natural intentar que el intervalo de ajuste de la ley sea lo más extenso posible [10]. Para determinar el intervalo se ha considerado una partición S caracterizada por la relación

$$\left\{ \forall s_i \in S, \frac{s_i}{s_{i-1}} = \sqrt[30]{10} \right\}$$

o dicho de otro modo, la partición divide cada década en 30 partes iguales. Sean $A, B \in S$ con $A < B$. Como se ha comentado anteriormente, maximizando la función de verosimilitud se obtiene α_d y mediante un contraste de bondad de ajuste se puede valorar si el ajuste de la ley de potencias en el intervalo propuesto es aceptable o no.

Finalmente, de entre todos los intervalos $[A_i, B_i]$ aceptables, se escoge el intervalo $[A, B]$ que cumpla $\frac{B}{A} = \max_{A_i, B_i \in S} \frac{B_i}{A_i}$.

Un intervalo será considerado aceptable en caso que el p-valor obtenido en el contraste de bondad de ajuste no sea inferior al 30 %.

2.6. Contraste de Bondad de Ajuste de Kolmogorov-Smirnov

El contraste de Bondad de ajuste utilizado para determinar el intervalo de ajuste $[A, B]$ es el contraste de Kolmogorov-Smirnov (KS) [13][19]:

$$\begin{cases} H_0 : F(x) = \frac{1 - \left(\frac{A}{x}\right)^{\alpha_d - 1}}{1 - r^{\alpha_d - 1}} \\ H_1 : F(x) \neq \frac{1 - \left(\frac{A}{x}\right)^{\alpha_d - 1}}{1 - r^{\alpha_d - 1}} \end{cases} \quad (2.18)$$

Donde el estadístico de Kolmogorov-Smirnov es

$$\mathcal{D} \equiv \sup_{A \leq x \leq B} |F_d(x) - F(x)| \quad (2.19)$$

siendo $F_d(x)$ la distribución empírica de los datos, es decir, la aproximación a la función de distribución definida a partir de los datos disponibles:

$$F_d(x) = \begin{cases} 0 & \text{si } x < x_1 \\ \frac{i}{N_{AB}} & \text{si } x_i \leq x < x_{i+1} \text{ con } i \in \{1, \dots, N_{AB} - 1\} \\ 1 & \text{si } x_{N_{AB}} \leq x \end{cases} \quad (2.20)$$

y $x_1 < x_2 < \dots < x_{N_{AB}}$ son los datos ordenados.

Para decidir la aceptación del ajuste hay que recurrir al p-valor del estadístico \mathcal{D} ya que éste es la probabilidad de no aceptar la hipótesis nula siendo cierta, o sea la probabilidad de cometer un error de tipo I. Si bien el p-valor puede ser obtenido directamente en R mediante la instrucción `ks.test` [3], éste no se puede utilizar ya que la distribución sometida al contraste es precisamente la que se ha construido a partir de los datos para que sea la que mejor se ajusta a ellos. Para evitar este problema se calcula el p-valor mediante el Método de Montecarlo [22]. La idea consiste en estimar la probabilidad de que el estadístico \mathcal{D}_{sim} que se puede obtener de una muestra de la ley de potencias que se ha ajustado sea mayor o igual que \mathcal{D} , utilizando muestras aleatorias de la ley de potencias.

2.7. Cálculo del p-valor de KS

Sea X el conjunto total de datos, $[A, B]$ el intervalo de ajuste de la ley de potencias, $N_{AB} = \#\{x \in X | A \leq x \leq B\}$, α_d el exponente máximo verosímil de la ley de potencias ajustada al intervalo $[A, B]$ y \mathcal{D} el estadístico asociado al contraste de Kolmogorov-Smirnov. El cálculo del p-valor del contraste [9] mediante Montecarlo consiste en simular k muestras aleatorias de longitud N_{AB} bajo la hipótesis nula y ajustar una ley de potencias a cada una de ellas obteniendo a su vez un estadístico de Kolmogorov-Smirnov \mathcal{D}_{sim} . El p-valor, que se define como la probabilidad de que el estadístico sea mayor o igual que \mathcal{D} si la hipótesis nula es cierta, se puede estimar mediante la expresión

$$p = \frac{\text{Número de simulaciones con } \mathcal{D}_{\text{sim}} \geq \mathcal{D}}{\text{Número de simulaciones}} \quad (2.21)$$

El número v de simulaciones cumpliendo $\mathcal{D}_{\text{sim}} \geq \mathcal{D}$ es una variable aleatoria binomial $B(k, p)$ y por tanto tiene desviación típica $\sigma_v = \sqrt{kp(1-p)}$. De acuerdo a la ecuación 2.21, $p = \frac{v}{k}$ y por tanto $\sigma_p = \frac{\sigma_v}{k} = \sqrt{\frac{p(1-p)}{k}}$ [11].

σ_p permite concretar el número de simulaciones a realizar si se pretende limitar la incertidumbre del p-valor siempre y cuando el valor de p sea intermedio. En el presente trabajo, el cálculo del p-valor y su desviación se realiza mediante $k = 1000$ simulaciones.

Sean $u_1, \dots, u_{N_{AB}}$ los N_{AB} valores simulados de una variable aleatoria uniforme en $[0, 1]$. Cada muestra aleatoria $Y = \{y_1, \dots, y_{N_{AB}}\}$ se construye simulando la ley de potencias mediante la fórmula

$$y = \frac{A}{[1 - (1 - r^{\alpha_d-1})u]^{1/(\alpha_d-1)}} \quad (2.22)$$

que se puede deducir a partir de la función de distribución $F(x)$ imponiendo $F(y) = u$ siendo u el valor de una variable aleatoria uniforme en $[0, 1]$:

$$\begin{aligned} F(y) = u \Rightarrow \frac{y^{1-\alpha_d} - A^{1-\alpha_d}}{B^{1-\alpha_d} - A^{1-\alpha_d}} = u \Rightarrow y^{1-\alpha_d} = u(B^{1-\alpha_d} - A^{1-\alpha_d}) + A^{1-\alpha_d} \Rightarrow \\ y = \frac{1}{[u(B^{1-\alpha_d} - A^{1-\alpha_d}) + A^{1-\alpha_d}]^{1/(\alpha_d-1)}} = \frac{A}{[u(r^{\alpha_d-1} - 1) + 1]^{1/(\alpha_d-1)}} \end{aligned}$$

2.8. Comparación de las distribuciones de probabilidad

Sean α_{PDI} , α_{OEI} y α_{IEI} los exponentes de las leyes de potencias para cada índice en una de las zonas de estudio. Considero ahora el siguiente contraste de hipótesis:

$$\begin{cases} H_0 : \alpha_{\text{PDI}} = \alpha_{\text{OEI}} = \alpha_{\text{IEI}} \\ H_1 : \text{No hay igualdad entre los 3 exponentes} \end{cases}$$

Tal y como se muestra en [5], el exponente estimado de la ley de potencias sigue una distribución normal de media α_d y varianza $\sigma_{\alpha_d}^2$.

Considero ahora los intervalos al 95 % de confianza de cada uno de los exponentes de las distribuciones ajustadas a los PDI, los IEI y a los OEI en cada una de las zonas (ver figura 3.4).

El contraste de hipótesis puede llevarse a cabo a partir de la intersección de los intervalos de confianza: en caso de que el intervalo obtenido en la intersección de los tres intervalos de confianza para los exponentes sea no vacío, se concluye la aceptación de la hipótesis nula (se acepta H_0). Si en cambio el

intervalo de intersección es vacío, no se acepta la hipótesis nula.
En caso de no aceptar H_0 , continuaré con la comparación entre las distribuciones de los índices pero de dos en dos. Dicho de otro modo consideraré los contrastes de la forma

$$\begin{cases} H_0 : \alpha_1 = \alpha_2 \\ H_1 : \alpha_1 \neq \alpha_2 \end{cases}$$

siendo α_1 y α_2 los exponentes de las leyes de potencias comparadas.

3 Resultados

Distribuciones de probabilidad ajustadas

Los resultados obtenidos en el ajuste de las leyes de potencias (ver cuadro 3.1) indican que el PDI ($m^3 s^{-2}$) en el norte del Atlántico se ajusta a

$$D(x) = 3.692x^{-1.125} \text{ si } x \in [1.08 \cdot 10^9, 1.47 \cdot 10^{11}] \quad (3.1)$$

mientras que en el noreste del Pacífico,

$$D(x) = 79.034x^{-1.252} \text{ si } x \in [1.47 \cdot 10^9, 10^{11}] \quad (3.2)$$

Los resultados en los ajustes obtenidos para el ajuste del IEI ($m^5 s^{-2}$) (cuadro 3.2) son

$$D(x) = 1.34 \cdot 10^{-2}x^{-0.938} \text{ si } x \in [3.41 \cdot 10^{18}, 3.16 \cdot 10^{20}] \quad (3.3)$$

en el caso del norte Atlántico, mientras que para el noreste del Pacífico:

$$D(x) = 3.3 \cdot 10^{11}x^{-1.615} \text{ si } x \in [9.26 \cdot 10^{18}, 2.15 \cdot 10^{20}] \quad (3.4)$$

Finalmente, los resultados para los ajustes del OEI ($m^5 s^{-2}$) (cuadro 3.3) han sido

$$D(x) = 1.3x^{-1.044} \text{ si } x \in [6.31 \cdot 10^{19}, 4.64 \cdot 10^{22}] \quad (3.5)$$

$$D(x) = 1.681x^{-1.047} \text{ si } x \in [5.41 \cdot 10^{19}, 1.71 \cdot 10^{22}] \quad (3.6)$$

para el norte del Océano Atlántico y noreste del Océano Pacífico respectivamente.

Destaca el hecho de que las estimaciones de la distribución del PDI son consistentes con las obtenidas en [10] por Corral, Ossó y Llebot, utilizando

otras bases de datos. Un intervalo de confianza del 95 % para el exponente del ajuste obtenido por Corral et al. en el norte del Atlántico es [1.0724, 1.3076], mientras que en el este del Pacífico el intervalo es [1.077, 1.273]. Al compararlos con los obtenidos aquí (ver cuadro 3.4), acepto coincidencia.

Comparativamente, la ley de potencias estimada para el OEI se ajusta mejor que la obtenida en el caso de IEI en el sentido que el rango de ajuste $\frac{B}{A}$ es mayor (ver figura 3.2). La similitud entre las expresiones que definen ambos índices hace creer que su comportamiento será similar, pero hay una diferencia fundamental que afecta a la caracterización de estos índices así como al proceso de completación de datos: se ha supuesto que las funciones que definen los radios respecto del tiempo son funciones continuas.

En el caso del radio de la isobara exterior cerrada del huracán la evolución es ciertamente continua mientras que para el caso del radio de máxima velocidad del viento, no necesariamente es así. Puede suceder que haya un radio distinto del de velocidad máxima cuya velocidad del viento iguale a la máxima velocidad en un determinado instante y a continuación la supere convirtiéndose en el nuevo radio de velocidad máxima. Si se produce este fenómeno de discontinuidad en la función radio de velocidad máxima, su comportamiento es más impredecible que en el caso de la función radio de isobara exterior cerrada. Como consecuencia, la compleción de los datos en el caso de los radios de máxima velocidad será, en general, menos eficiente que la de los radios de isobara exterior cerrada.

Comparación entre las distribuciones de probabilidad

En el caso del Océano Atlántico la intersección de los intervalos de confianza (cuadro 3.4) de cada uno de los exponentes de las leyes de potencias es el intervalo [1.0425, 1.0829], por tanto se puede aceptar la hipótesis de que las tres leyes de potencias tienen el mismo exponente. Dicho de otro modo, se acepta equivalencia entre los distintos índices. Para el Océano Pacífico la situación es un poco distinta ya que la intersección de los tres intervalos de confianza es vacía. Si se intersecan los intervalos de confianza de dos en dos (o lo que es lo mismo, se procede a decidir sobre los contrastes de hipótesis que comparan pares de distribuciones), se puede aceptar que todas las parejas de exponentes son iguales excepto en el caso de $D(OEI)$ y $D(IEI)$.

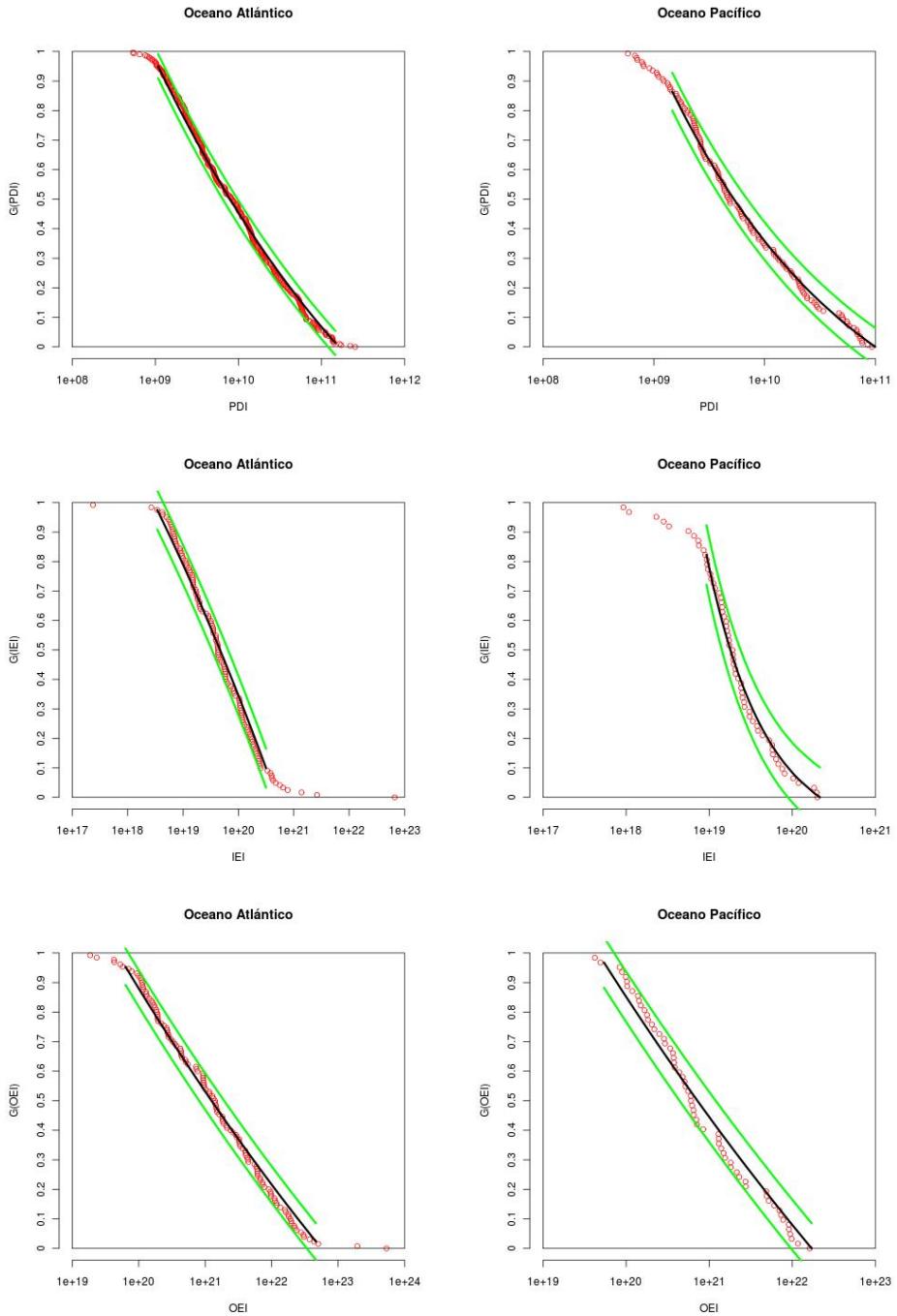


Figura 3.1: Representación simultánea de las funciones complementarias a las funciones de distribución empíricas y las obtenidas a partir del ajuste de la ley de potencias. La líneas verdes corresponden a la distancia de Kolmogorov-Smirnov (el estadístico \mathcal{D}).

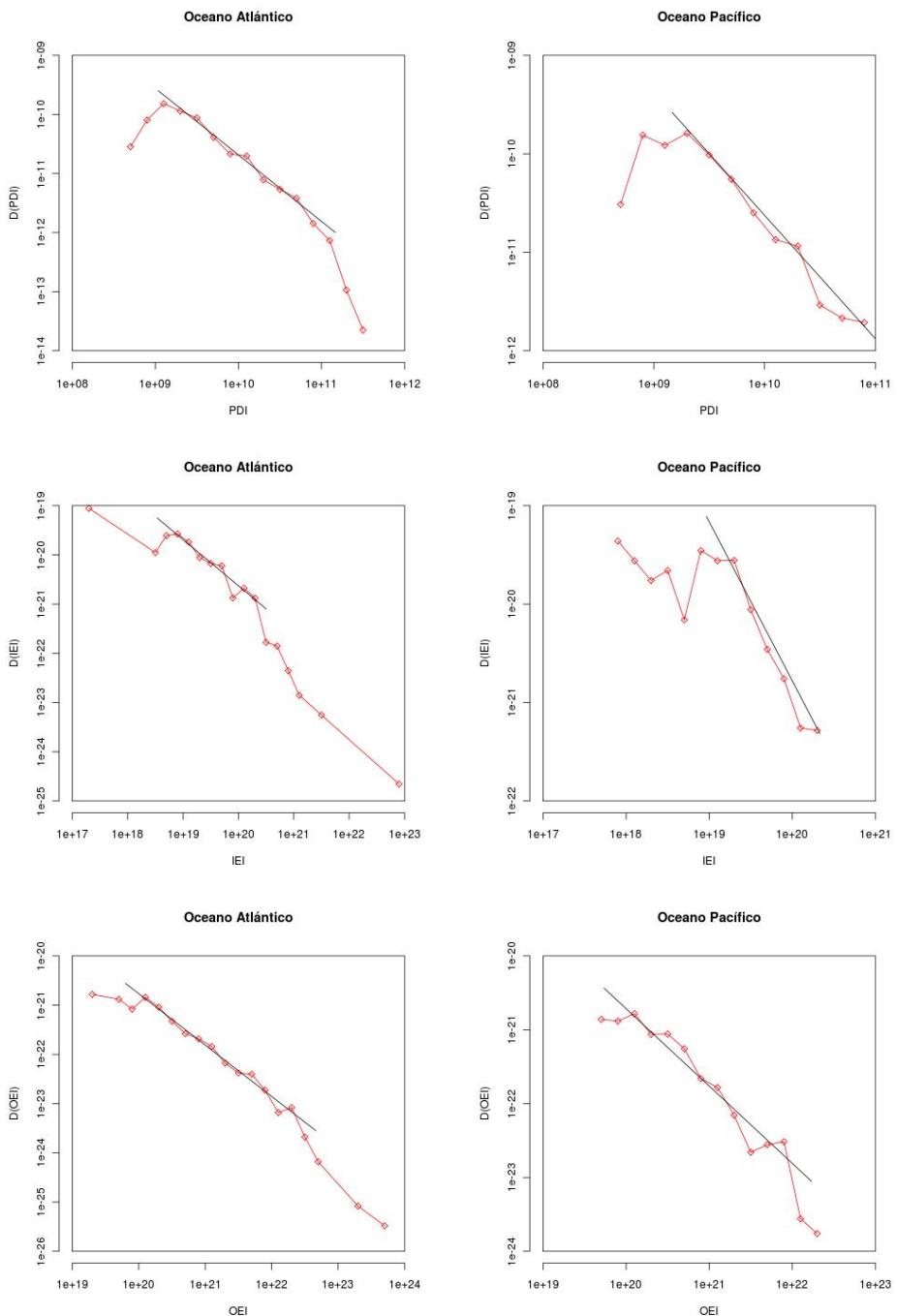


Figura 3.2: Representación simultánea de las funciones de densidad empíricas aproximadas y las obtenidas a partir del ajuste de la ley de potencias.

	Océano Atlántico	Océano Pacífico
N	303	140
A	$1.08 \cdot 10^9$	$1.47 \cdot 10^9$
B	$1.47 \cdot 10^{11}$	$1 \cdot 10^{11}$
N_{AB}	284	121
α_d	1.125	1.252
σ_{α_d}	$4.22 \cdot 10^{-2}$	$7.67 \cdot 10^{-2}$
C	3.692	79.034
\mathcal{D}	$4.096 \cdot 10^{-2}$	$6.393 \cdot 10^{-2}$
p	0.341	0.311
σ_p	$1.499 \cdot 10^{-2}$	$1.464 \cdot 10^{-2}$

Cuadro 3.1: Resultados obtenidos en el ajuste de sendas leyes de potencias para los datos del norte del Océano Atlántico y del noreste del Pacífico. N es el número total de ciclones con PDI calculado. A y B (en $m^3 s^{-2}$) son los extremos del intervalo de ajuste óptimo. N_{AB} es el número de ciclones cuyo PDI está comprendido entre A y B . α_d es el exponente de la ley de potencias ajustada, cuya desviación típica es σ_{α_d} . C es la constante que normaliza la ley de potencias en el intervalo $[A, B]$. \mathcal{D} corresponde al valor del estadístico de Kolmogorov-Smirnov del ajuste y p se refiere al p-valor para A y B fijados. Finalmente, σ_p es la desviación típica del p-valor.

	Océano Atlántico	Océano Pacífico
N	120	57
A	$3.41 \cdot 10^{18}$	$9.26 \cdot 10^{18}$
B	$3.16 \cdot 10^{20}$	$2.15 \cdot 10^{20}$
N_{AB}	107	51
α_d	0.938	1.615
σ_{α_d}	$7.41 \cdot 10^{-2}$	$1.68 \cdot 10^{-1}$
C	$1.338 \cdot 10^{-2}$	$3.3 \cdot 10^{11}$
\mathcal{D}	$6.628 \cdot 10^{-2}$	$1.013 \cdot 10^{-1}$
p	0.328	0.305
σ_p	$1.485 \cdot 10^{-2}$	$1.456 \cdot 10^{-2}$

Cuadro 3.2: Del mismo modo que en el cuadro 3.1, en esta tabla se recogen los resultados correspondientes al ajuste del IEI. En este caso A y B en $m^5 s^{-2}$.

	Océano Atlántico	Océano Pacífico
N	128	62
A	$6.31 \cdot 10^{19}$	$5.41 \cdot 10^{19}$
B	$4.64 \cdot 10^{22}$	$1.71 \cdot 10^{22}$
N_{AB}	121	60
α_d	1.044	1.047
σ_{α_d}	$4.78 \cdot 10^{-2}$	$7.78 \cdot 10^{-2}$
C	1.3	1.68
\mathcal{D}	$6.170 \cdot 10^{-2}$	$8.511 \cdot 10^{-2}$
p	0.325	0.346
σ_p	$1.459 \cdot 10^{-2}$	$1.504 \cdot 10^{-2}$

Cuadro 3.3: Resultados correspondientes al ajuste del OEI en cada una de las dos zonas. Como con el IEI, A y B en $m^5 s^{-2}$.

	Atlántico		Pacífico	
PDI	1.0425	1.2081	1.1019	1.4026
OEI	0.9503	1.1377	0.8945	1.1996
IEI	0.7925	1.0829	1.2847	1.9450

Cuadro 3.4: Intervalos al 95 % de confianza de los exponentes de las leyes de potencias ajustadas del PDI, el OEI y el IEI en las dos zonas de estudio.

4 Conclusiones

La base de datos contenido la información del noreste del Océano Pacífico es demasiado pequeña como para obtener resultados fiables. De todos modos es remarcable el acuerdo en los resultados con los obtenidos por Corral et al. utilizando distintas bases de datos, datos correspondientes a años distintos y con nuevos programas que han sido desarrollados específicamente para este proyecto.

El ajuste de las distribuciones de probabilidad del tipo ley de potencias del PDI, el OEI y el IEI no basta para considerar la posibilidad de que no se produzca una influencia significativa del calentamiento global en las intensidades de los huracanes que se tienen lugar en el noreste del Océano Pacífico y en el norte del Océano Atlántico. Para poder valorar esa posibilidad sería conveniente realizar un estudio detallado por años. Esta es una opción de estudio en un futuro inmediato.

La falta de certeza en la continuidad de la función radio de máxima velocidad del viento hace que al aproximar los valores de los índices en tiempo discreto sea aconsejable valorar la energía del huracán mediante el OEI cuyos resultados, junto con los del PDI, han sido satisfactorios (a pesar de aparente sesgo en la selección de datos).

Al comparar la disipación (PDI) con la energía (OEI) se puede aceptar que ambas distribuciones de probabilidad tienen el mismo exponente y por tanto se acepta equivalencia entre ambas.

El método de compleción de las bases de datos abre otra opción de trabajo futuro: un estudio sistemático que determine el porcentaje de compleción asumible para minimizar el sesgo en la selección de datos.

Bibliografía

- [1] RMMB, Regional and Mesoscale Meteorology Branch.
<http://rammb.cira.colostate.edu/>
- [2] NOAA, National Oceanic and Atmospheric Administration.
<http://www.noaa.gov/>
- [3] The R Project for Statistical Computing. <http://www.r-project.org/>
- [4] NHC, National Hurricane Center. <http://www.nhc.noaa.gov/>
- [5] I. B. Aban, M. M. Meerschaert, A. K. Panorska. Parameter Estimation for the Truncated Pareto Distribution. *Journal of the American Statistical Association*. Vol. 101, No. 473, *Theory and Methods*. 270-277 (2006)
- [6] C. Arthur, A. Schofield, B. Cechet. Assessing the impacts of tropical cyclones. *The Australian Journal of Emergency Management*, Vol. 23 No. 4, 14-20 (2008).
- [7] P. Bak. How Nature Works: The Science of Self-Organised Criticality. *Copernicus Press, New York* (1996).
- [8] R. L. Burden, J. D. Faires. Análisis numérico. Segunda Edición. *Grupo Editorial Iberoamérica, S.A. de C.V.* (1996)
- [9] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51, 661-703 (2009)
- [10] A. Corral, A. Ossó, J. E. Llebot. Scaling of tropical-cyclone dissipation. *Nature Physics* vol 6, 693-696 (2010).

- [11] A. Corral. Dependence of earthquake recurrence times and independence of magnitudes on seismicity history. *Tectonophysics* 424, 177-193 (2006).
- [12] A. Corral, A. Turiel. Variability of North Atlantic Hurricanes: Seasonal Versus Individual-Event Features. <http://hdl.handle.net/2072/196868>. (2011).
- [13] C. M. Cuadras. Problemas de Probabilidades y Estadística. Volumen 2: Inferencia Estadística. *Promociones y Publicaciones Universitarias, S.A.* (1991)
- [14] K. Emanuel. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature Vol.436*, 686-688 (2005).
- [15] K. Emanuel. The power of a hurricane: An example of reckless driving on the information superhighway. *Weather* 54, 107-108 (1999).
- [16] A. Graumann et al. Hurricane Katrina, A Climatological Perspective. *NOAA's National Climatic Data Center* (2005-2006).
- [17] S. Milojević. Power law Distributions in Information Science: Making the Case for Logarithmic Binning. *Journal of the American Society for Information Science and Technology*. Vol. 61, No.12, 2417-2425 (2010).
- [18] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46, 323-351 (2005).
- [19] V. Novo. Estadística Teórica y Aplicada. *Editorial Sanz y Torres* (2004)
- [20] T. Schott, C. Landsea y otros. The Saffir-Simpson Hurricane Wind Scale. *NOAA* (2012).
- [21] F. Seychelles et al. Thermal Convection and Emergence of Isolated Vortices in Soap Bubbles. *Phys. Rev. Lett. Vol 100, issue 14* (2008).
- [22] I. M. Sóbol. Método de Montecarlo. *Lecciones Populares de Matemáticas*. *Editorial MIR* (1983).
- [23] E. Strobl. The Economic Growth Impact of Hurricanes: Evidence from US Coastal Counties. *Ecole Polytechnique Paris and IZA. Discussion Paper No. 3619* (2008).
- [24] H. E. Willoughby et al. Project STORMFURY: A Scientific Chronicle 1962-1983. *Hurricane Research Division. AOML/NOAA* (1985)

Parte II

Anexo

1 Relación entre el OEI, el IEI y la energía de un ciclón tropical

En un fenómeno ciclónico, la cantidad de energía cinética generada es igual a la que se disipa por la fricción. La potencia disipada en cualquier instante se estima por medio de la fórmula [12] [15]

$$P(t) = \int \rho C_D \nu^3 d^2r$$

donde t es el tiempo, r se refiere a las coordenadas sobre la superficie terrestre, $\rho = \rho(r, t)$ es la densidad del aire, $C_D = C_D(r, t)$ es el coeficiente de fricción y $\nu = \nu(r, t)$ corresponde al módulo de la velocidad del viento.

Esta fórmula puede simplificarse. En [15] Emanuel tomó como valores representativos para la densidad del aire y el coeficiente de fricción, $\rho = 1 \text{ Kg m}^{-3}$ y $C_D = 2 \cdot 10^{-3}$. Además consideró un sencillo perfil funcional para ν :

$$\nu(r, t) = v_m(t) f\left(\frac{r}{R(t)}\right)$$

con $v_m(t)$ como la máxima velocidad del viento dentro del fenómeno en el instante t , $R(t)$ algún tipo de radio característico del ciclón y f una función de escala común para todos los ciclones.

Se tiene,

$$P(t) = \int \rho C_D \nu^3 d^2r \approx \rho C_D v_m^3(t) \int f^3\left(\frac{r}{R(t)}\right) d^2r =$$

considérese ahora el cambio de variable $u = \frac{r}{R(t)}$, obteniéndose

$$P(t) \approx \hat{P}(t) = \rho C_D v_m^3(t) R^2(t) \int f^3(u) d^2u \propto v_m^3(t) R^2(t)$$

se concluye pues que

$$E \approx \int \hat{P}(t) dt \propto \int v_m^3(t) R^2(t) dt$$

y esta última expresión corresponde al IEI si $R(t)$ es el radio de máxima velocidad del viento y al OEI si es el radio de la isobara exterior.

Conviene destacar que la constante de proporcionalidad no es de interés para el propósito del presente trabajo.

2 Código fuente en R

- *velocidad* calcula la velocidad en $m \cdot s^{-1}$ a partir de la velocidad en nudos.

```
velocidad=function(kt){  
  return(0.514444444*kt)}
```

- *distancia* calcula longitudes en metros a partir de las longitudes en millas náuticas.

```
distancia=function(nm){  
  return(1852*nm)}
```

- *MLF.eq* devuelve el exponente máximo verosímil, su desviación típica y su verosimilitud. *A* y *B* son los límites del intervalo de ajuste. *l* es la lista de valores del ajuste. *ainf* y *asup* son los límites entre los que se determina el exponente *a*.

```
MLF.eq=function(A,B,l,n,ainf,asup){  
  l=l[common(l>=A,l<=B)];  
  l=l[sort.list(l)];  
  N=length(l);  
  r=A/B;  
  M=maximization(A,B,l,n,ainf,asup);  
  a=M[1];  
  ver=M[2];  
  s2=var.alpha(a,r,N);  
  return(c(a,sqrt(s2),ver))}
```

- *var.alpha* calcula la varianza del exponente *a*. *r* es la razón de los extremos del intervalo de ajuste y *N* es el número de datos utilizados para estimar el exponente.

```

var.alpha=function(a,r,N){
  return((1/(1/(a-1)^2-
  (r^(a-1))*(log(r)^2)/((1-r^(a-1))^2)))/N)}

```

- *maximization* maximiza el logaritmo de la función de verosimilitud. A y B son los límites del intervalo de ajuste, l es la lista de valores del ajuste, a_{inf} y a_{sup} son los límites entre los que se maximiza la función. Esta función devuelve el máximo y su imagen para el logaritmo de la función de verosimilitud.

```

maximization=function(A,B,l,a_{inf},a_{sup}){
  r=A/B;
  v=common(l<=B,l>=A);
  l=l[v];
  N=length(l);
  F=function(x){
    return(log((x-1)/(1-r^(x-1))-
    (x/N)*sum(log(l))+(x-1)*log(A))};
  values=optimize(F,interval=c(a_{inf},a_{sup}),maximum=TRUE,
  tol=1e-8);
  return(c(values$maximum,values$objective))}

```

- *common* es una función muy simple que devuelve un vector lógico indicando los elementos comunes de los vectores v y w .

```

common=function(v,w){
  return(as.logical(v*w))}

```

- *fit* es la función principal y calcula el ajuste de la ley de potencias. *LISTA* es la lista de valores a ajustar, *a_{inf}* y *a_{sup}* son los valores entre los que se busca el exponente de la ley de potencias, *min.PV* es el valor mínimo que ha de tener el p-valor de un ajuste para ser aceptado y *particion* es el número de intervalos en que se divide cada década en la partición para determinar el intervalo de ajuste.

```

fit=function(LISTA,a_{inf},a_{sup},min.PV=0.3,particion=30){
  LISTA=LISTA[sort.list(LISTA)];
  I1=10^(floor(min(log10(LISTA))));
  I2=max(LISTA);
  S.LISTA=c(I1);
  while(S.LISTA[length(S.LISTA)]<I2){

```

```

S.LISTA=c(S.LISTA,S.LISTA[length(S.LISTA)]*10^(1/particion));
V=S.LISTA<=min(LISTA);
S.LISTA=S.LISTA[S.LISTA>=max(S.LISTA[V])];
D=1e10;
RANGE=0;
PV=-1;
P.sigma=0;
ListA=S.LISTA;
ListB=S.LISTA;
M=length(LISTA);
NewA=0;
NewB=0;
alpha=0;
sigma=0;
ver=0;
casos=0;
for(j in 1:(length(ListA)-1)){
A=ListA[j];
v2=LISTA>=A;
for(k in (j+1):length(ListB)){
B=ListB[k];
print(c(A,B,j,k));
v1=LISTA<=B;
v=common(v1,v2);
N=sum(v);
if(N>10){
R=B/A;
values=MLF.eq(A,B,LISTA[v],100,ainf,asup);
actual=values[1];
sigma.actual=values[2];
ver.actual=values[3];
D.actual=fit.ks.2(LISTA,A,B,actual);
if(R>RANGE){
Nsim=1000;
PV.actual=MCsim(A,B,ainf,asup,actual,LISTA,Nsim);
if(PV.actual>0){
P.sigma.actual=PV.actual*sqrt((1-PV.actual)/(PV.actual*Nsim));
else{P.sigma.actual=0};
if(PV.actual>min.PV){
D=D.actual;
NewA=A;
}
}
}
}
}
}

```

```

NewB=B;
PV=PV.actual;
P.sigma=P.sigma.actual;
alpha=actual;
sigma=sigma.actual;
ver=ver.actual;
casos=N;RANGE=R}}}}}};
return(c(NewA,NewB,alpha,sigma,D,PV,P.sigma,casos))}
```

- *fit.2* es una función auxiliar que calcula el ajuste de las muestras aleatorias generadas de la ley de potencias.
LISTA es el vector de valores a ajustar. *a_{inf}* y *a_{sup}* son los límites entre los que determinar el exponente del ajuste. *A* y *B* son los extremos del intervalo de ajuste de la ley de potencias a partir de los datos empíricos.

```

fit.2=function(LISTA,ainf,asup,A,B){
LISTA=LISTA[sort.list(LISTA)];
D=1e10;
S=c(A,B);
ListA=S;
ListB=S;
M=length(LISTA);
for(j in 1:(length(ListA)-1)){
A=ListA[j];
v2=LISTA>=A;
for(k in j:length(ListB)){
B=ListB[k];
v1=LISTA<=B;
v=common(v1,v2);
N=sum(v);
if((B/A)>20){
values=MLF.eq(A,B,LISTA[v],100,ainf,asup);
actual=values[1];
sigma.actual=values[2];
ver.actual=values[3];
D.actual=fit.ks.2(LISTA,A,B,actual);
if(D.actual<D){
D=D.actual;
NewA=A;
NewB=B;
alpha=actual;
```

```

sigma=sigma.actual} } } } ;
return(D)

```

- *fit.ks.2* es otra función auxiliar. Ésta calcula el estadístico de Kolmogorov-Smirnov para las muestras aleatorias generadas de la ley de potencias. *LISTA* es el vector de datos ajustados. *A* y *B* son los extremos del intervalo de ajuste de la ley de potencias a partir de los datos empíricos. *a* es el exponente estimado para el ajuste.

```

fit.ks.2=function(LISTA,A,B,a){
v1=LISTA<=B;
v2=LISTA>=A;
w=sum(LISTA<A);
N=length(LISTA);
v=common(v1,v2);
F=function(x){return((x^(1-a)-A^(1-a))/(B^(1-a)-A^(1-a)))};
test=ks.test(LISTA[v],F,alternative="two.sided");
D=test$statistic;
return(c(D))
}

```

- *power.law* es una función que genera muestras aleatorias de la ley de potencias de exponente *a* y definida entre *A* y *B* a partir de valores de una variable aleatoria uniforme en [0,1].

```

power.law=function(A,B,a){
r=A/B;
return(A/(1-(1-r^(a-1))*runif(1))^(1/(a-1)))}

```

- *MCsim* calcula el p-valor del ajuste de la ley de potencias en $[A, B]$ con exponente *a* mediante montecarlo. *ainf* y *asup* son los límites entre los que determinar el exponente del ajuste de las muestras aleatorias. *LISTA* es el vector de datos empíricos. *n* es el número de simulaciones a efectuar.

```

MCsim=function(A,B,ainf,asup,a,LISTA,n){
N=length(LISTA);
v1=LISTA<=B;
v2=LISTA>=A;
v=common(v1,v2);
NAB=sum(v);
dvalues=c();

```

```

D=fit.ks.2(LISTA,A,B,a);
m=0;
for(i in 1:n){
  data=c();
  for(j in 1:NAB){
    data=c(data,power.law(A,B,a))};
  dvalues=c(dvalues,fit.2(data,ainf,asup,A,B))};
  return(sum(dvalues>=D)/n)}

```

- *constante* calcula la constante de normalización de la ley de potencias ajustada a $[A, B]$ con exponente a .

```

constante=function(a,A,B){
r=A/B;
return(((a-1)*A^(a-1))/(1-r^(a-1)))}

```

- *splines* interpola y extrapola utilizando trazadores cúbicos. $X1$ es el vector de abscisas de los puntos a interpolar. $Y1$ es el vector de ordenadas de los puntos a interpolar. $Z1$ es el vector de abscisas a evaluar en la interpolación.

```

splines=function(X1,Y1,Z1){
X=X1[sort.list(X1)];
Y=Y1[sort.list(X1)];
N=length(Y);
COEF=matrix(rep(0,4*N),ncol=4);
COEF[,1]=Y[1:N];
H=c();
for(i in 1:(N-1)){H=c(H,X[i+1]-X[i])};
ALFA=c(0);
for(i in 2:(N-1)){
  ALFA=c(ALFA,3*(Y[i+1]-Y[i])/H[i]-3*(Y[i]-Y[i-1])/H[i-1])};
l=c(1);
mu=c(0);
z=c(0);
for(i in 2:(N-1)){
  l=c(l,2*(X[i+1]-X[i-1])-H[i-1]*mu[i-1]);
  mu=c(mu,H[i]/l[i]);
  z=c(z,(ALFA[i]-H[i-1]*z[i-1])/l[i])};
l=c(l,1);
z=c(z,0);

```

```

for(j in (N-1):1){
COEF[j,3]=z[j]-mu[j]*COEF[j+1,3];
COEF[j,2]=(COEF[j+1,1]-COEF[j,1])/H[j]-
H[j]*(COEF[j+1,3]+2*COEF[j,3])/3;
COEF[j,4]=(COEF[j+1,3]-COEF[j,3])/(3*H[j])};
evaluacion=function(v,L,xj){
s=L[1];
for(i in 2:length(L)){s=s+L[i]*(v-xj)^(i-1)};
return(s)};
R=c();
for(i in 1:length(Z1)){
if(Z1[i]<X[1]){R=c(R,evaluacion(Z1[i],COEF[1,],X[1]))};
if(Z1[i]>X[N]){R=c(R,evaluacion(Z1[i],COEF[N-1,],X[N-1]))};
if(Z1[i]>=X[1] && Z1[i]<=X[N]){
A=max(which(X<=Z1[i]));
R=c(R,evaluacion(Z1[i],COEF[A,],X[A]))};
return(R)}

```

- *unificar* concatena los vectores contenidos en la lista de vectores L .

```

unificar=function(L){
R=c();
N=length(L);
for(i in 1:N){
for(j in 1:length(L[[i]])){R=c(R,L[[i]][j])}};
return(R)}

```

- *grafico.densidad* representa la estimación de la función de densidad de una lista de valores mediante logarithmic binning.
 $LISTA$ es la lista total de valores a tener en cuenta para aproximar la función densidad. *titulo*, *ejex* y *ejey* son el título del gráfico, el del eje X y el del eje Y , respectivamente.
Si *ajuste=TRUE*, se representa el ajuste de la ley de potencias entre A y B con exponente a .

```

grafico.densidad=function(LISTA,titulo=" ",ejex=" ",
ejey=" ",ajuste=FALSE,A=0,B=0,a=0){
I1=10^(floor(min(log10(LISTA))));
I2=max(LISTA);
W=c(I1);
while(W[length(W)]<I2){

```

```

W=c(W,W[length(W)]*10^(1/5));
V=W<=min(LISTA);
W=W[W>=max(W[V])];
W2=c();
D=c();
for(i in 2:length(W)){
W2=c(W2,sqrt(W[i-1]*W[i]));
v=common(LISTA>W[i-1],LISTA<=W[i]);
D=c(D,sum(v)/(W[i]-W[i-1])));
D=D/length(LISTA);
for(i in 1:length(D)){if(D[i]==0){D[i]=NA}};
X0=floor(min(log10(LISTA)));
XF=ceiling(max(log10(LISTA)));
Y0=floor(min(log10(D[is.na(D)==FALSE])));
YF=ceiling(max(log10(D[is.na(D)==FALSE])));
if(ajuste==FALSE){
grafico=plot(log10(W2[is.na(D)==FALSE]),
log10(D[is.na(D)==FALSE]),main=titulo,
xlab=ejex,ylab=ejey,axes=FALSE,type="o",col="red",
pch=23,xlim=c(X0,XF),ylim=c(Y0,YF));
axis(1,at=c(X0:XF),labels=10^c(X0:XF));
axis(2,at=c(Y0:YF),labels=10^c(Y0:YF));
points(c(X0,X0,XF,XF,X0),c(Y0,YF,YF,Y0,Y0),type="l");
if(ajuste==TRUE){
intercept=log10(constante(a,A,B));
grafico=plot(log10(W2[is.na(D)==FALSE]),
log10(D[is.na(D)==FALSE]),main=titulo,
xlab=ejex,ylab=ejey,axes=FALSE,type="o",col="red",
pch=23,xlim=c(X0,XF),ylim=c(Y0,YF));
axis(1,at=c(X0:XF),labels=10^c(X0:XF));
axis(2,at=c(Y0:YF),labels=10^c(Y0:YF));
points(c(X0,X0,XF,XF,X0),c(Y0,YF,YF,Y0,Y0),type="l");
lines(c(log10(A),log10(B)),
c(intercept-a*log10(A),intercept-a*log10(B)))};
return(grafico)
}

```

- *grafico.ajuste* representa gráficamente la función complementaria a la función de distribución empírica, así como la ajustada.
LISTA es la lista total de valores a representar. $[A, B]$ es el intervalo de ajuste. a es el exponente obtenido en el ajuste. D es el estadístico de Kolmogorov-Smirnov del ajuste. *titulo*, *ejex* y *ejey* son el título del

gráfico, el del eje X y el del eje Y, respectivamente.

```
grafico.ajuste=
function(LISTA,A,B,a,D,titulo=" ",ejex=" ",ejey=" "){
x=c(A,LISTA[common(LISTA<=B,LISTA>=A)],B);
N=length(LISTA);
y=1-c(1:N)/N;
k1=sum(LISTA<A);
k2=sum(LISTA>B);
F=function(x,A,B,a){
return(k1/N+
((N-k1-k2)/N)*(x^(1-a)-A^(1-a))/(B^(1-a)-A^(1-a)));
X0=floor(min(log10(LISTA)));
XF=ceiling(max(log10(LISTA)));
grafico=plot(log10(LISTA),y,col="red",main=titulo,
xlab=ejex,ylab=ejey,axes=FALSE,xlim=c(X0,XF),ylim=c(0,1));
axis(1,at=c(X0:XF),labels=10^c(X0:XF));
axis(2,at=c(0:10)/10,labels=c(0:10)/10);
points(c(X0,X0,XF,XF,X0),c(0,1,1,0,0),type="l");
lines(log10(x),1-F(x,A,B,a),type="l",lwd=3);
lines(log10(x),1-F(x,A,B,a)+D,col="green",lwd=3);
lines(log10(x),1-F(x,A,B,a)-D,col="green",lwd=3);
return(grafico)}
```

- *imputacion* completa los radios que faltan en la base de datos indicada en función de los criterios de restricción que se indiquen.

TABLA es la base de datos (*Atlántico* o *Pacífico*) a usar. *prop.max* es la proporción máxima admitida en la imputación de datos faltantes para los radios de máxima velocidad. *prop.iso* es la proporción máxima admitida en la imputación de datos faltantes para los radios de las isobaras exteriores.

Esta rutina devuelve los vectores con los nuevos radios así como vectores *W* y *V* indicando los huracanes imputados.

```
imputacion=function(TABLA,prop.max=0.2,prop.iso=0.2){
N=length(TABLA[,1])
L1=c(1);
k=1;
for(i in 2:N){
if(TABLA$name[i]!=TABLA$name[i-1]){k=k+1};L1=c(L1,k)}
na=L1[length(L1)];
```

```

RADIOS.MAX=list();
RADIOS.ISO=list();
for(i in 1:na){
  RADIOS.MAX[[i]]=c(TABLA$rad.max.speed[L1==i]);
  RADIOS.ISO[[i]]=c(TABLA$out.isobar.rad[L1==i])};
  ratio.miss=c(rep(0,na));
  for(i in 1:na){
    ratio.miss[i]=
    sum(RADIOS.MAX[[i]]== -99)/length(RADIOS.MAX[[i]])}
    isobar.miss=c(rep(0,na));
    for(i in 1:na){
      isobar.miss[i]=
      sum(RADIOS.ISO[[i]]== -99)/length(RADIOS.ISO[[i]])}
      W2=ratio.miss<prop.max;
      W=which(W2==TRUE);
      V2=isobar.miss<prop.iso;
      V=which(V2==TRUE);
      SPLINES.RAD.MAX=RADIOS.MAX;
      SPLINES.RAD.ISO=RADIOS.ISO;
      for(j in 1:length(W)){
        T=RADIOS.MAX[[W[j]]] != -99;
        SPLINES.RAD.MAX[[W[j]]]=
        splines(which(T==TRUE),
        RADIOS.MAX[[W[j]]][which(T==TRUE)],
        c(1:length(RADIOS.MAX[[W[j]]])))}
        for(j in 1:length(V)){
          T=RADIOS.ISO[[V[j]]] != -99;
          SPLINES.RAD.ISO[[V[j]]]=
          splines(which(T==TRUE),
          RADIOS.ISO[[V[j]]][which(T==TRUE)],
          c(1:length(RADIOS.ISO[[V[j]]])))}
          for(i in 1:na){
            n1=length(SPLINES.RAD.MAX[[i]]);
            n2=length(SPLINES.RAD.ISO[[i]]);
            for(j in 1:n1){
              SPLINES.RAD.MAX[[i]][j]=max(0,SPLINES.RAD.MAX[[i]][j])};
              for(j in 1:n2){
                SPLINES.RAD.ISO[[i]][j]=max(0,SPLINES.RAD.ISO[[i]][j])}};
                R.MAX=unificar(SPLINES.RAD.MAX);
                R.ISO=unificar(SPLINES.RAD.ISO);
                return(list(R.MAX,R.ISO,W,V))}
```

- *calculo.valores* calcula el PDI, IEI y OEI de la base de datos indicada. *TABLA* es la base de datos (*Atlántico* o *Pacífico*) a usar. *R.MAX* es la lista de radios de máxima velocidad. *R.ISO* es la lista de radios de la isobara exterior. *W* es el vector con los huracanes a considerar en el cálculo de IEI. *V* es el vector con los huracanes a considerar en el cálculo de OEI. *R.MAX*, *R.ISO*, *W* y *V* se pueden obtener a partir de la función *imputacion*.

Esta rutina devuelve tres vectores; el primero contiene los PDI para todos los huracanes, el segundo los IEI para los huracanes indicados en *W* y el tercero los OEI para los huracanes indicados en *V*.

```
calculo.valores=function(TABLA,R.MAX,R.ISO,W,V){
N=length(TABLA[,1]);
L1=c(1);
k=1;
for(i in 2:N){
  if(TABLA$name[i]!=TABLA$name[i-1]){k=k+1};L1=c(L1,k)}
na=L1[length(L1)];
PDI=c(rep(0,na));
IEI=c(rep(0,na));
OEI=c(rep(0,na));
for(i in 1:N){
  PDI[L1[i]]=PDI[L1[i]]+
  6*3600*velocidad(TABLA$max.speed[i])^3;
  IEI[L1[i]]=IEI[L1[i]]+
  (velocidad(TABLA$max.speed[i])^3)*
  (distancia(R.MAX[i])^2)*6*3600;
  OEI[L1[i]]=OEI[L1[i]]+
  (velocidad(TABLA$max.speed[i])^3)*
  (distancia(R.ISO[i])^2)*6*3600};
PDI=sort.list(PDI);
IEI=IEI[W];
OEI=OEI[V];
IEI=sort.list(IEI);
OEI=sort.list(OEI);
return(list(PDI,IEI,OEI))}
```