

# Modelización estadística de extremos de velocidad de viento



**Eduardo Benedí Belmonte**  
Trabajo de fin de grado de Matemáticas  
Universidad de Zaragoza

Directores del trabajo:  
Jesús Asín Lafuente Y Zeus Gracia Tabuenca  
19 de Enero de 2023





# Summary

The modeling and prediction of extreme values of geophysical variables, such as wind, ocean surface waves, sea level, temperature and river flow, has always been a field of main concern for engineers and scientists. The analysis of extreme wind speed particularly plays an important role in natural disasters' preparedness, prevention, mitigation and management and in various ocean, environmental and civil engineering applications, such as the design of offshore platforms and coastal marine structures, coastal management, wind climate analysis.

In fact, in the special supplement of the Bulletin of the American Meteorological Society, “*State of the Climate in 2022*” [5], in Figure Plate 1.1, there are global (or representative) average time series on the evolution of different essential climatic variables, among them, the surface wind velocity, which is the one that interests us to start the approach of our work.

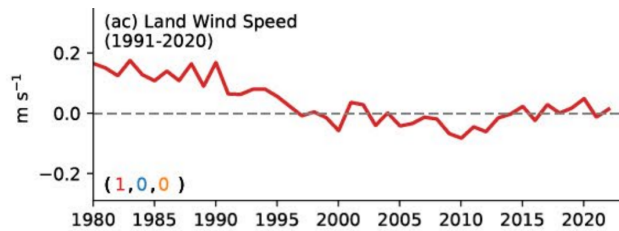


Figure 1: Annual global mean annual onshore wind speed for the period 1991 – 2020 measured in meters per second. Anomalies are shown relative to the base period in parentheses. The numbers in parentheses at the bottom left indicate how many in situ (red), reanalysis (blue) and satellite (orange) data sets are used to create the time series in that order.

The figure shows the global average wind speed in each year at about 1500 meters altitude, and we can see that over the years it decreases. This decrease suggests that the winds are slower, so there is evidence that it is decreasing in the world, and reinforces the idea of non-stationarity in the behavior of this. Most studies on wind variability have focused on calm versus reversal phenomena at global and regional scales, while long-term changes in winds at local scales, such as the sea breeze (SB onwards), represent a knowledge gap in climate research.

In this work we are going to rely on a study of wind variability, more specifically on SBs at coastal stations called “*Opposite trends of sea-breeze speeds and gusts in Eastern Spain, 1961–2019*” [2] and detailed in Annex A.B. In this study we have developed a robust automated method for the identification of SB days, and by means of homogenized wind observations from 16 stations in eastern Spain, we analyze the multidecadal variability and trends in speeds, gusts and SB occurrence for 1961 – 2019. The main finding is the opposite trends and uncoupled variability of SB speeds and gusts.

## 0.1. Database

A brief description of the data available is presented in order to perform an exploratory analysis necessary to understand wind behavior in each zone.

The dataset we are going to work with contains information of the maximum speed series recorded daily in the observatories of the study from 1961 to 2019. It consists of a `data.frame` composed of a column where we have the date, the following ones correspond to each observatory, which is named with the AEMET indicator for the corresponding place (shown in Annex A.C), and finally, three columns indicating the year, month and day of the year corresponding to each record respectively.



Figura 2: Own-authored map with the 16 weather stations that make up the dataset.

## 0.2. Objective

Climate change is enhanced by extreme events, and a thorough analysis of the local effects will allow us to check if there really is a direct relationship with wind decrease, which is the objective of this work. For this purpose, we will use the extreme value analysis (EVA) R package **extRemes** version 2.0, whose functions mainly provide tools to implement univariate EVA, with a focus on meteorological and climatic applications, including the incorporation of covariates.

To do this, we will need to review the theoretical results of extreme values, and with the objective of checking whether or not there are changes, we will also review the extremes of non-stationary sequences. Then, we will perform an exploratory analysis of the data we have at various stations, whose results will provide us with the way to build an optimal model, that will allow us to statistically check with the quantiles and return levels if there is indeed a decrease in wind over the years.

# Índice general

<b>Summary</b>	<b>III</b>
0.1. Database . . . . .	III
0.2. Objective . . . . .	IV
<b>1. Introducción</b>	<b>1</b>
1.1. Base de datos . . . . .	2
1.2. Objetivo . . . . .	2
<b>2. Fundamentos teóricos</b>	<b>3</b>
2.1. Teoría y modelos clásicos del valor extremo . . . . .	3
2.1.1. Modelos asintóticos . . . . .	3
2.1.2. Inferencia para la distribución GEV . . . . .	6
2.2. Extremos de secuencias no estacionarias . . . . .	7
2.2.1. Estructuras del modelo . . . . .	7
2.2.2. Inferencia . . . . .	8
<b>3. Metodología propuesta para el análisis de la evolución de la velocidad del viento</b>	<b>11</b>
3.1. Análisis exploratorio . . . . .	11
3.2. Modelización inicial . . . . .	12
3.3. Modelización óptima . . . . .	12
3.4. extRemes 2.0: un paquete de análisis de valor extremo . . . . .	13
<b>4. Resultados</b>	<b>15</b>
4.1. Análisis exploratorio de los datos . . . . .	15
4.1.1. Aeropuerto de Reus . . . . .	15
4.1.2. Costas Nororiental y Este . . . . .	16
4.2. Modelización inicial . . . . .	17
4.2.1. Aeropuerto de Reus . . . . .	18
4.3. Modelización óptima . . . . .	19
4.3.1. Aeropuerto de Reus . . . . .	20
4.3.2. Otros observatorios . . . . .	21
<b>5. Conclusiones</b>	<b>23</b>
<b>Bibliografía</b>	<b>27</b>
<b>ANEXO</b>	<b>29</b>
A.A. Conceptos teóricos . . . . .	31
A.A.1. Modelos asintóticos para mínimos . . . . .	31
A.B. Tendencias opuestas de las velocidades y ráfagas de las brisas marinas en el este de España, 1961-2019 . . . . .	33

A.B.1. Introducción . . . . .	33
A.B.2. Datos y métodos . . . . .	33
A.B.3. Resultados . . . . .	36
A.B.4. Discusión . . . . .	38
A.B.5. Conclusiones . . . . .	39
A.C. Indicadores AEMET . . . . .	41
A.D. Análisis exploratorio . . . . .	43
A.D.1. Aeropuerto de Reus . . . . .	43
A.D.2. Gráficas de la velocidad máxima y desviación típica por zonas . . . . .	46
A.E. Modelización inicial . . . . .	49
A.E.1. Aeropuerto de Reus . . . . .	49
A.E.2. Aeropuerto de Almería . . . . .	50
A.E.3. Estudio regional . . . . .	51
A.E.4. Valores de $\mu_1$ . . . . .	52
A.F. Modelización óptima . . . . .	53
A.F.1. Aeropuerto de Reus . . . . .	53
A.F.2. Aeropuerto de Almería . . . . .	53
A.F.3. Barcelona Fabra . . . . .	60
A.F.4. Tablas de valores AIC . . . . .	65
A.G. Conclusiones . . . . .	69
A.G.1. Tablas informativas de cuantiles . . . . .	69
A.G.2. Gráficas de los cambios absolutos y relativos en las tres estaciones . . . . .	70

# Capítulo 1

## Introducción

La modelización y predicción de valores extremos de variables geofísicas, como el viento, las olas de la superficie del océano, el nivel del mar, la temperatura y el caudal de los ríos, ha sido siempre un campo de gran interés para ingenieros y científicos. En particular, el análisis de la velocidad extrema del viento desempeña un papel importante en la preparación, prevención, mitigación y gestión de catástrofes naturales, así como en diversas aplicaciones oceánicas, medioambientales y de ingeniería civil, como el diseño de plataformas marinas y estructuras marinas costeras, la gestión del litoral, el análisis del régimen eólico.

De hecho, en el suplemento especial del Boletín de la Sociedad Meteorológica Americana, “*State of the Climate in 2022*” [5], en la figura **Plate 1.1** aparecen series temporales medias mundiales (o representativas) sobre la evolución de diferentes variables climáticas, entre ellas, la velocidad del viento en superficie, que es la que interesa para iniciar el planteamiento de nuestro trabajo.

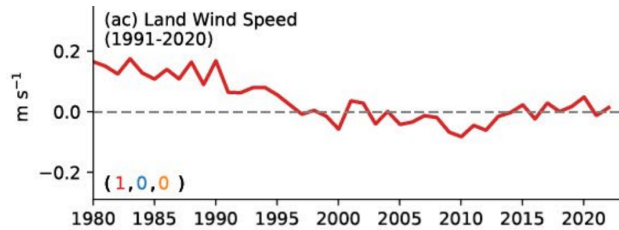


Figura 1.1: Velocidad media global anual del viento en superficie terrestre correspondiente a 1991 – 2020, en metros por segundo. Las anomalías se muestran en relación con el periodo base, entre paréntesis. Los números entre paréntesis en la parte inferior izquierda indican cuántos conjuntos de datos in situ (rojo), de reanálisis (azul) y de satélite (naranja) se utilizan para crear la serie temporal en ese orden.

La figura muestra el valor medio global de la velocidad del viento en cada año, podemos observar que a lo largo de los años disminuye. Esta disminución da a entender que los vientos son ahora más lentos, luego hay evidencia de que está decayendo y se refuerza la idea de no estacionariedad en su comportamiento. La mayoría de los estudios sobre la variabilidad del viento han profundizado en los fenómenos de calma versus inversión a escala global y regional, mientras que los cambios a largo plazo en los vientos a escala local, como la brisa marina (SB en adelante), representan una brecha de conocimiento en la investigación climática.

En este trabajo vamos a basarnos en un estudio de variabilidad del viento, más concretamente en las estaciones costeras, llamado “*Opposite trends of sea-breeze speeds and gusts in Eastern Spain, 1961–2019*” [2] y detallado en el Anexo A.B. En este estudio se ha desarrollado un método automatizado robusto para la identificación de días SB, y mediante las observaciones de viento homogeneizadas de 16 estaciones en el este de España, se analiza la variabilidad multidecenal

y las tendencias en las velocidades, ráfagas y ocurrencia de SB para 1961 – 2019. El hallazgo principal son las tendencias opuestas y variabilidad desacoplada de las velocidades y ráfagas de SB.

## 1.1. Base de datos

Se presenta una breve descripción de los datos que disponemos para poder realizar un análisis exploratorio necesario para entender los comportamientos del viento en cada zona.

El conjunto de datos con el que vamos a trabajar contiene información de las series de velocidad máxima registrada diariamente en los observatorios del estudio desde 1961 hasta 2019. Consta de un `data.frame` compuesto por una columna donde tenemos la fecha, las siguientes corresponden a cada observatorio, que se denomina con el indicador de AEMET para el lugar correspondiente (mostrados en el Anexo A.C), y por último, tres columnas indicando el año, el mes y el día del año que corresponde a cada registro respectivamente.



Figura 1.2: Posición de las 16 estaciones meteorológicas.

## 1.2. Objetivo

El cambio climático está potenciado por los eventos extremos, y un análisis exhaustivo de los efectos locales nos va a permitir comprobar si realmente hay una relación directa con la disminución de viento, objetivo de este trabajo. Para ello, usaremos el paquete R de análisis de valores extremos (EVA) **extRemes** versión 2.0, cuyas funciones proporcionan principalmente utilidades para estimar modelos EVA, con un enfoque en aplicaciones meteorológicas y climáticas, incluida la incorporación de covariables.

Para ello, necesitaremos revisar los fundamentos teóricos de valores extremos y, con ese objetivo de comprobar si hay o no cambios, también se revisarán los extremos de secuencias no estacionarias. Seguidamente, realizaremos un análisis exploratorio de los datos que disponemos en diversas estaciones, cuyos resultados nos facilitarán la forma de construir un modelo óptimo, que nos permitirá comprobar estadísticamente con los cuantiles y niveles de retorno si en efecto hay una disminución de viento con el paso de los años.

## Capítulo 2

# Fundamentos teóricos

Se revisan los conceptos de teoría y modelos clásicos del valor extremo, incluyendo modelos asintóticos, para mínimos (Anexo A.A), e inferencia para la distribución valor extremo generalizada (GEV). Además, se revisan resultados para extremos de secuencias no estacionarias.

### 2.1. Teoría y modelos clásicos del valor extremo

#### 2.1.1. Modelos asintóticos

##### Formulación del modelo

Vamos a desarrollar el modelo que representa la piedra angular de la teoría del valor extremo, centrado en el comportamiento estadístico de

$$M_n = \max\{X_1, \dots, X_n\},$$

donde  $X_1, \dots, X_n$  es una secuencia de variables aleatorias independientes que tienen una función de distribución común  $F$ . Notar que las  $X_i$  representan valores de un proceso medidos en una escala temporal regular, como por ejemplo las mediciones máximas anuales en un punto de costa del nivel del mar. Además,  $M_n$  representa el máximo del proceso en  $n$  observaciones en un año.

La distribución de  $M_n$  deducida para los valores de  $n$  viene dada por la función  $F_n(z)$ :

$$F_n(z) = Pr\{M_n \leq z\} = Pr\{X_1 \leq z, \dots, X_n \leq z\} = Pr\{X_1 \leq z\} \cdot \dots \cdot Pr\{X_n \leq z\} = \{F(z)\}^n.$$

Esta expresión no es útil porque  $F$  no es conocida. Además, aunque podemos recurrir a técnicas estadísticas estándar para estimar  $F$  a partir de los datos observados, tenemos que discrepancias muy pequeñas en la estimación de  $F$  pueden dar lugar a discrepancias sustanciales para  $F_n(z)$ .

Un enfoque alternativo consiste en aceptar que  $F$  es desconocida y buscar familias aproximadas de modelos para  $F^n$ , que puedan estimarse únicamente a partir de los datos extremos. Procedemos observando el comportamiento de  $F^n$  a medida que  $n \rightarrow \infty$ . Pero esto por sí solo no es suficiente: para cualquier  $z < z_+$ , donde  $z_+$  es el punto extremo superior de  $F$ ,<sup>1</sup> tenemos que  $F^n(z) \rightarrow 0$  a medida que  $n \rightarrow \infty$ , de modo que la distribución de  $M_n$  degenera en una masa puntual en  $z_+$ . Esta dificultad se evita permitiendo una renormalización lineal de la variable  $M_n$ :

$$M_n^* = \frac{M_n - b_n}{a_n},$$

para secuencias de constantes  $\{a_n > 0\}$  y  $\{b_n\}$ . Las elecciones apropiadas de  $\{a_n\}$  y  $\{b_n\}$  estabilizan la localización y escala de  $M_n^*$  a medida que  $n$  aumenta, evitando las dificultades que surgen con la variable  $M_n$ . Por lo tanto, buscamos distribuciones límite para  $M_n^*$ , con elecciones apropiadas de  $\{a_n\}$  y  $\{b_n\}$ , en lugar de  $M_n$ .

---

<sup>1</sup> $z_+$  es el menor valor de  $z$  tal que  $F(z) = 1$ .

### Teorema de los tipos extremos

Toda la gama de posibles distribuciones límite para  $M_n^*$  viene dada por el **Teorema 3.1**, el teorema de los tipos extremos, mostrado en el libro “*An Introduction to Statistical Modeling of Extreme Values*” [1].

**Teorema 3.1** Si existen secuencias de constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tales que

$$Pr \left\{ \frac{M_n - b_n}{a_n} \leq z \right\} \longrightarrow G(z) \quad \text{cuando } n \longrightarrow \infty,$$

donde  $G$  es una función de distribución no degenerada, entonces  $G$  pertenece a una de las siguientes familias:

$$\begin{aligned} \text{I:} \quad & G(z) = \exp \left\{ -\exp \left[ -\left( \frac{z-b}{a} \right) \right] \right\}, \quad -\infty < z < \infty; \\ \text{II:} \quad & G(z) = \begin{cases} 0, & z \leq b, \\ \exp \left\{ -\left( \frac{z-b}{a} \right)^{-\alpha} \right\}, & z > b; \end{cases} \\ \text{III:} \quad & G(z) = \begin{cases} \exp \left\{ -\left[ -\left( \frac{z-b}{a} \right)^\alpha \right] \right\}, & z < b, \\ 1, & z \geq b, \end{cases} \end{aligned}$$

para los parámetros  $a > 0$ ,  $b$  y, en el caso de las familias **II** y **III**,  $\alpha > 0$ .  $\square$

Estas tres clases de distribución se denominan **distribuciones de valor extremo**, y los tipos **I**, **II** y **III** se conocen como familias de **Gumbel**, **Fréchet** y **Weibull**, respectivamente. Cada familia tiene un parámetro de localización y de escala,  $b$  y  $a$  respectivamente; además, las familias Fréchet y Weibull tienen un parámetro de forma  $\alpha$ . El **Teorema 3.1** implica que, cuando  $M_n$  puede estabilizarse con secuencias  $\{a_n\}$  y  $\{b_n\}$  adecuadas, la variable normalizada correspondiente  $M_n^*$  tiene una distribución límite que debe ser uno de los tres tipos de distribución de valor extremo. La característica notable de este resultado es que los tres tipos de distribuciones de valores extremos son los únicos límites posibles para las distribuciones de la  $M_n^*$ , independientemente de la distribución  $F$  de la población.

### La distribución de valor extremo generalizada

Los tres tipos de límites que surgen en el **Teorema 3.1** tienen formas distintas de comportamiento en las colas, dependiendo de la función de distribución  $F$  de la  $X_i$ . Esto se puede precisar considerando el comportamiento de la distribución límite  $G$  en  $z_+$ , su extremo superior. Para la distribución de Weibull  $z_+$  es finito, mientras que para las distribuciones de Fréchet y Gumbel  $z_+ = \infty$ . Sin embargo, la densidad de  $G$  decae exponencialmente para la distribución de Gumbel y polinómicamente para la distribución de Fréchet, respondiendo a tasas relativamente diferentes de decaimiento en la cola de  $F$ .

Una reformulación de los modelos del **Teorema 3.1** ofrece un mejor análisis. Las familias de Gumbel, Fréchet y Weibull pueden combinarse en una única familia de modelos con funciones de distribución

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right)^{-1/\xi} \right] \right\}, \quad (1)$$

definida en el conjunto  $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$ , donde los parámetros satisfacen  $-\infty < \mu < \infty$ ,  $\sigma > 0$  y  $-\infty < \xi < \infty$ . Se trata de la familia de distribuciones de **valor extremo generalizado** (GEV). El modelo tiene tres parámetros: un parámetro de localización  $\mu$ , un parámetro de escala  $\sigma$ , y un parámetro de forma  $\xi$ . Las clases tipo **II** y tipo **III** de la distribución de valores extremos corresponden respectivamente a los casos  $\xi > 0$  y  $\xi < 0$  en esta parametrización. El subconjunto



de la familia GEV con  $\xi = 0$  se interpreta como el límite de (1) como  $\xi \rightarrow 0$ , lo que conduce a la familia de Gumbel con función de distribución

$$G(z) = \exp \left[ -\exp \left\{ -\left( \frac{z - \mu}{\sigma} \right) \right\} \right], \quad -\infty < z < \infty.$$

La unificación de las tres familias originales de distribución de valores extremos en una única familia simplifica enormemente la aplicación estadística. A través de la inferencia sobre  $\xi$ , los propios datos determinan el tipo más apropiado de comportamiento de la cola, y no hay necesidad de seleccionar sobre qué familia individual de valores extremos adoptar. Además, la incertidumbre en el valor inferido de  $\xi$  mide la falta de certeza sobre cuál de los tres tipos originales es el más apropiado para un conjunto de datos determinado. Por comodidad, volvemos a exponer el **Teorema 3.1** de forma modificada.

**Teorema 3.1.1** Si existen secuencias de constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tales que

$$\Pr \left\{ \frac{M_n - b_n}{a_n} \leq z \right\} \rightarrow G(z) \quad \text{cuando } n \rightarrow \infty,$$

para una función de distribución no degenerada  $G$ , entonces  $G$  es un miembro de la familia GEV

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right)^{-1/\xi} \right] \right\},$$

definida en  $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$ , donde  $-\infty < \mu < \infty$ ,  $\sigma > 0$  y  $-\infty < \xi < \infty$ .  $\square$

Si el **Teorema 3.1.1** permite la aproximación de la distribución de  $M_n^*$  por un miembro de la familia GEV para  $n$  grande, la distribución de la propia  $M_n$  también puede ser aproximada por un miembro diferente de la misma familia.

Este argumento conduce al siguiente enfoque para modelizar los extremos de una serie de observaciones independientes  $X_1, X_2, \dots$ . Los datos se bloquean en secuencias de observaciones de longitud  $n$ , para algún valor grande de  $n$ , generando una serie de  $m$  máximos de bloque,  $M_{n,1}, \dots, M_{n,m}$ , a los que se puede ajustar la distribución GEV. A menudo, los bloques se eligen para que correspondan a un período de tiempo de un año, en cuyo caso  $n$  es el número de observaciones de un año y los máximos de los bloques son máximos anuales. Las estimaciones de los cuantiles extremos de la distribución de máximos anuales se obtienen invirtiendo la ecuación (1):

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{-\log(1-p)\}^{-\xi}], & \text{para } \xi \neq 0, \\ \mu - \sigma \log\{-\log(1-p)\}, & \text{para } \xi = 0, \end{cases} \quad (2)$$

donde  $G(z_p) = 1 - p$ .

Tenemos que  $z_p$  es el **nivel de retorno** asociado al **periodo de retorno**  $1/p$ , ya que con un grado razonable de precisión, se espera que el nivel  $z_p$  se supere por término medio una vez cada  $1/p$  años. Más concretamente,  $z_p$  es superado por el máximo anual en cualquier año concreto con probabilidad  $p$ . Dado que los cuantiles permiten expresar los modelos probabilísticos a escala de los datos, definiendo  $y_p = -\log(1-p)$ , tenemos que

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - y_p^{-\xi}], & \text{para } \xi \neq 0, \\ \mu - \sigma \log y_p, & \text{para } \xi = 0, \end{cases}$$

se deduce que, si  $z_p$  se representa gráficamente frente a  $y_p$  en una escala logarítmica (o, lo que es lo mismo, si  $z_p$  se representa gráficamente frente a  $\log y_p$ ), el gráfico es lineal en el caso  $\xi = 0$ . Si  $\xi < 0$ , el gráfico es convexo con límite asintótico  $p \rightarrow 0$  en  $\mu - \sigma/\xi$ ; si  $\xi > 0$ , el gráfico es cóncavo y no tiene límite finito. Este gráfico es un **gráfico de nivel de retorno**. Debido a la sencillez de su interpretación y a que la elección de la escala comprime la cola de la distribución de forma que se pone de relieve el efecto de la extrapolación, los gráficos de nivel de retorno son especialmente convenientes tanto para la presentación como para la validación de modelos. La figura 2.1 muestra gráficos de niveles de retorno para tres ejemplos de  $\xi$ .

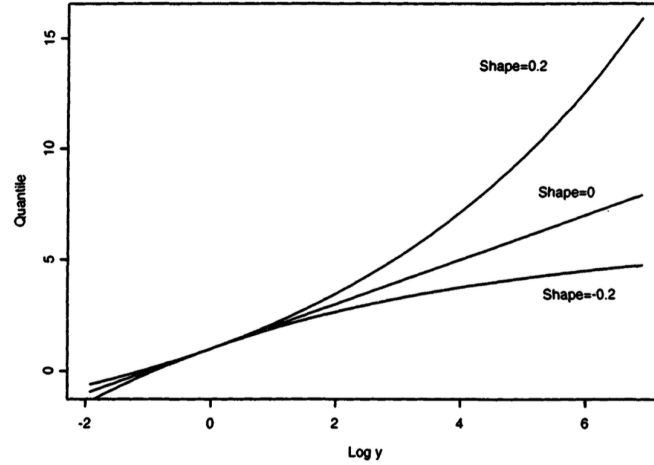


Figura 2.1: Gráficos de nivel de retorno de la distribución GEV con parámetros de forma  $\xi = -0.2$ ,  $\xi = 0$  y  $\xi = 0.2$  respectivamente.

### 2.1.2. Inferencia para la distribución GEV

Del **Teorema 3.1.1**, la GEV proporciona un modelo para la distribución de los máximos de bloque. Su aplicación consiste en resumir los datos en bloques de igual longitud y ajustar la GEV al conjunto de máximos de bloque. Sin embargo, a la hora de aplicar este modelo a un conjunto de datos concreto, la elección del tamaño de los bloques puede ser fundamental.

Los parámetros  $\mu$ ,  $\sigma$  y  $\xi$  se estiman por máxima verosimilitud. Conocemos que la estimación de  $z_p$ , como función de  $\mu$ ,  $\sigma$  y  $\xi$  es la función aplicada sobre las estimaciones ML de  $\hat{\mu}$ ,  $\hat{\sigma}$  y  $\hat{\xi}$ . Mediante la sustitución de las estimaciones de máxima verosimilitud de los parámetros GEV en (2), la estimación de máxima verosimilitud de  $z_p$  para  $0 < p < 1$ , el nivel de retorno  $1/p$ , se obtiene como

$$\hat{z}_p = \begin{cases} \mu - \frac{\hat{\sigma}}{\hat{\xi}}[1 - y_p^{-\hat{\xi}}], & \text{para } \hat{\xi} \neq 0, \\ \hat{\mu} - \hat{\sigma} \log y_p, & \text{para } \hat{\xi} = 0, \end{cases}$$

donde  $y_p = -\log(1 - p)$ . Además, por el método delta,

$$\text{Var}(\hat{z}_p) \approx \nabla z_p^T V \nabla z_p, \quad (3)$$

donde  $V$  es la matriz de varianza-covarianza de  $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$  y

$$\nabla z_p^T = \left[ \frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \sigma}, \frac{\partial z_p}{\partial \xi} \right] = [1, -\xi^{-1}(1 - y_p^{-\xi}), \sigma \xi^{-2}(1 - y_p^{-\xi}) - \sigma \xi^{-1} y_p^{-\xi} \log y_p]$$

evaluada en  $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ .

Suelen ser los periodos de retorno largos, correspondientes a valores pequeños de  $p$ , los de mayor interés. Si  $\hat{\xi} < 0$ , también es posible hacer inferencias sobre el extremo superior de la distribución, que es efectivamente el "periodo de retorno de observación infinito", correspondiente a  $z_p$  con  $p = 0$ . La estimación de máxima verosimilitud es

$$\hat{z}_0 = \hat{\mu} - \hat{\sigma}/\hat{\xi}$$

y (3) sigue siendo válida con

$$\nabla z_0^T = [1, -\xi^{-1}, \sigma \xi^{-2}]$$

evaluada de nuevo en  $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ . Cuando  $\hat{\xi} \geq 0$  la estimación de máxima verosimilitud del punto final superior es infinito. Dada la variabilidad, es necesario ser cauteloso en la interpretación de las inferencias de los niveles de retorno muy elevados.

## 2.2. Extremos de secuencias no estacionarias

### 2.2.1. Estructuras del modelo

Los procesos no estacionarios tienen características que cambian sistemáticamente a lo largo del tiempo. En el contexto de los procesos medioambientales, la no estacionariedad suele manifestarse por efectos estacionales, quizá debidos a patrones climáticos diferentes en el ciclo verano-invierno, o en forma de tendencias, posiblemente debidas a cambios climáticos a largo plazo. Al igual que la presencia de dependencia temporal, tales desviaciones de los supuestos simples que se hicieron en la derivación de las caracterizaciones de valores extremos desafían la utilidad de los modelos estándar.

Sujetos a limitaciones específicas, los modelos habituales de valores extremos límite siguen siendo aplicables en presencia de dependencia temporal. No es posible establecer una teoría general de este tipo para los procesos no estacionarios. Se dispone de resultados para algunas formas muy especializadas de no estacionariedad, pero en general suelen ser demasiado restrictivos para describir los patrones de no estacionariedad que se dan en los procesos reales. En su lugar, es habitual adoptar un enfoque estadístico consistente en utilizar los modelos estándar de valores extremos mediante modelización estadística.

A modo de ejemplo, supongamos que tenemos unos datos de la velocidad máxima diaria de viento, con argumentos asintóticos que apoyan el uso de la distribución GEV para modelizar la velocidad máxima de viento en cualquier día, pero la tendencia aparente de los datos plantea dudas sobre la idoneidad de un modelo que supone una distribución constante a lo largo del tiempo debido a situaciones que implican estacionalidad como brisas nocturnas, periodo de verano, etc.

Existe una similitud entre la clase de modelos GEV y los modelos lineales generalizados (GLMs), cuya teoría está bien desarrollada y para los que se proporcionan algoritmos de estimación en el software estadístico. Sin embargo, la analogía no es lo suficientemente estrecha como para que los resultados estándar o las herramientas computacionales sean directamente transferibles al contexto de los valores extremos. La principal diferencia es que la familia de los GLM se limita a las distribuciones de la familia exponencial, mientras que los modelos estándar de valores extremos suelen quedar fuera de esta familia.

Utilizando la notación  $GEV(\mu, \sigma, \xi)$  para denotar la distribución GEV con parámetros  $\mu$ ,  $\sigma$  y  $\xi$ , se deduce que un modelo adecuado para  $Z_t$ , la velocidad máxima diaria de viento en el año  $t$ , podría ser

$$Z_t \sim GEV(\mu(t), \sigma, \xi),$$

donde

$$\mu(t) = \beta_0 + \beta_1 t$$

para los parámetros  $\beta_0$  y  $\beta_1$ . De este modo, las variaciones a lo largo del tiempo en el proceso observado se modelizan como una tendencia lineal en el parámetro de localización del modelo de valores extremos apropiado, que en este caso es la distribución GEV. El parámetro  $\beta_1$  corresponde a la tasa de variación anual de la velocidad máxima diaria de viento. El modelo homogéneo en el tiempo, que parece más plausible para las velocidades máximas diarias de viento, constituye un caso especial del modelo dependiente del tiempo, con  $\beta_1 = 0$ , en cuyo caso  $\mu = \beta_0$ .

También pueden ser apropiados cambios más complejos en  $\mu$ . Por ejemplo, un modelo cuadrático,

$$\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2,$$

o un modelo de punto de cambio,

$$\mu(t) = \begin{cases} \mu_1 & \text{para } t \leq t_0, \\ \mu_2 & \text{para } t > t_0. \end{cases}$$

La no estacionariedad también puede expresarse en términos de otros parámetros de valores extremos. Por ejemplo,

$$\sigma(t) = \exp(\beta_0 + \beta_1 t),$$

donde se utiliza la función exponencial para garantizar que se respeta la positividad de  $\sigma$  para todos los valores de  $t$ .

Otra situación que puede darse es que el comportamiento extremo de una serie esté relacionado con el de otra covariable, un vector de covariables  $X$ . Por ejemplo, en un análisis de valores extremos de concentraciones de contaminación, la extensión de un nivel alto de contaminante puede depender de la velocidad del viento concurrente (los vientos fuertes tienen un efecto dispersivo).

En cada caso, los parámetros de valor extremo pueden escribirse de la forma

$$\theta(X) = h(X^T \beta), \quad (4)$$

donde  $\theta$  denota  $\mu$ ,  $\sigma$  o  $\xi$  cualesquiera,  $h$  es una función especificada,  $\beta$  es un vector de parámetros y  $X$  es un vector de covariables. En este contexto,  $h$  suele denominarse función de enlace. Para la tendencia lineal en  $\mu$ ,  $h$  es la función de identidad y

$$\mu(t) = [1, t] \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}. \quad (5)$$

Esto se amplía a

$$\mu(t) = [1, t, t^2] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

para el modelo de tendencia cuadrática. El modelo log-lineal para  $\sigma$  tiene una estructura similar a (5), con  $\sigma(t)$  sustituyendo a  $\mu(t)$  y con el enlace inverso  $h$  tomado como función la exponencial. El modelo estacional con  $k$  estaciones  $s_1, \dots, s_k$  tiene la forma

$$\mu(t) = [I_1(t), I_2(t), \dots, I_k(t)] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix},$$

donde  $I_j(t)$  es la función indicadora

$$I_j(t) = \begin{cases} 1, & \text{si } s(t) = s_j, \\ 0, & \text{de lo contrario,} \end{cases}$$

y con expresiones similares para  $\sigma(t)$  y  $\xi(t)$ .

### 2.2.2. Inferencia

Una ventaja de la máxima verosimilitud sobre otras técnicas de estimación de parámetros es su adaptabilidad a los cambios en la estructura del modelo. Tomemos, por ejemplo, un modelo GEV no estacionario para describir la distribución de  $Z_t$  para  $t = 1, \dots, m$ :

$$Z_t \sim \text{GEV}(\mu(t), \sigma(t), \xi(t)),$$

donde  $\mu(t)$ ,  $\sigma(t)$  y  $\xi(t)$  tiene una expresión en términos de un vector de parámetros y covariables del tipo (4). Denotando por  $\beta$  el vector completo de parámetros, la verosimilitud es simplemente

$$L(\beta) = \prod_{t=1}^m g(z_t; \mu(t), \sigma(t), \xi(t)),$$

donde  $g(z_t; \mu(t), \sigma(t), \xi(t))$  denota la función de densidad GEV con parámetros  $\mu(t)$ ,  $\sigma(t)$ ,  $\xi(t)$  evaluados en  $z_t$ . Por lo tanto, si ninguno de los  $\xi(t)$  es cero, la log-verosimilitud es

$$\ell(\mu, \sigma, \xi) = - \sum_{t=1}^m \left\{ \log \sigma(t) + (1 + 1/\xi(t)) \log \left[ 1 + \xi(t) \left( \frac{z_t - \mu(t)}{\sigma(t)} \right) \right] + \left[ 1 + \xi(t) \left( \frac{z_t - \mu(t)}{\sigma(t)} \right) \right]^{-1/\xi(t)} \right\}, \quad (6)$$

siempre que

$$1 + \xi(t) \left( \frac{z_t - \mu(t)}{\sigma(t)} \right) > 0, \quad \text{para } t = 1, \dots, m,$$

donde  $\mu(t)$ ,  $\sigma(t)$  y  $\xi(t)$  se sustituyen por sus expresiones en términos de  $\beta$  según la estructura del modelo elegido determinada por (4). Si alguno de los  $\xi(t) = 0$ , es necesario utilizar la forma límite apropiada como  $\xi(t) \rightarrow 0$  en (6). Se pueden utilizar técnicas numéricas para maximizar (6), obteniendo la estimación de máxima verosimilitud de  $\beta$ . Los errores estándar y los intervalos de confianza aproximados se obtienen de la forma habitual a partir de la matriz de información observada, que también puede evaluarse numéricamente.

### Elección de modelo

Con la posibilidad de modelizar cualquier combinación de los parámetros del modelo de valores extremos como funciones del tiempo u otras covariables, existe un amplio catálogo de modelos entre los que elegir, y la selección de un modelo apropiado se convierte en una cuestión importante. El principio básico es la parsimonia, obtener el modelo más simple posible, que explique la mayor parte posible de la variación de los datos. Una opción podría ser necesitar un componente de tendencia lineal, probablemente en el parámetro de localización  $\mu$ . Pero, ¿quizás también podría ser evidente una tendencia cuadrática? Dado que la clase de modelos cuadráticos incluye los modelos lineales como un caso especial, un modelo de este tipo está destinado a mejorar el modelo lineal en términos de precisión en la descripción de las variaciones en los datos observados. Sin embargo, el modelo es necesario como descripción del proceso que generó los datos, no para los datos en sí, por lo que es necesario evaluar la solidez de las pruebas de la estructura más compleja del modelo. Si las pruebas no son especialmente sólidas, habrá que optar por el modelo más sencillo.

La estimación de máxima verosimilitud de modelos anidados conduce a un sencillo procedimiento de prueba de un modelo frente al otro. Con los modelos  $\mathcal{M}_0 \subset \mathcal{M}_1$ , el estadístico de desviación se define como

$$D = 2\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\},$$

donde  $\ell_0(\mathcal{M}_0)$  y  $\ell_1(\mathcal{M}_1)$  son las log-verosimilitudes maximizadas bajo los modelos  $\mathcal{M}_0$  y  $\mathcal{M}_1$  respectivamente. Los valores grandes de  $D$  indican que el modelo  $\mathcal{M}_1$  explica sustancialmente más de la variación en los datos que  $\mathcal{M}_0$ ; los valores pequeños de  $D$  sugieren que el aumento en el tamaño del modelo no trae mejoras valiosas en la capacidad del modelo para explicar los datos. La distribución asintótica de la función de desviación ayuda a determinar cómo de grande debe ser  $D$  para que el modelo  $\mathcal{M}_1$  sea preferible al modelo  $\mathcal{M}_0$ .

El modelo  $\mathcal{M}_0$  se rechaza mediante una prueba al nivel  $\alpha$  de significación si  $D > c_\alpha$ , donde  $c_\alpha$  es el cuantil  $(1 - \alpha)$  de la distribución  $\chi_k^2$ , y  $k$  es la diferencia en la dimensionalidad de  $\mathcal{M}_1$  y  $\mathcal{M}_0$ . En otras palabras, se puede dar un criterio formal para especificar cómo de grande debe ser  $D$  para que se prefiera el modelo  $\mathcal{M}_1$ . Sigue habiendo subjetividad en la elección de  $\alpha$ , cuyo valor determina el compromiso entre los dos tipos de error que se pueden cometer (adoptar erróneamente  $\mathcal{M}_1$ , o quedarse erróneamente con  $\mathcal{M}_0$ ), pero este acto de equilibrio es inevitable. Para los modelos no anidados, se han propuesto diversas modificaciones del criterio basado en la desviación (Cox & Hinkley, 1974, por ejemplo).

## Diagnóstico de modelos

Tras haber estimado una serie de modelos posibles y haber seleccionado entre ellos, debemos asegurarnos de que el modelo final adoptado sea realmente una buena representación de los datos. En el caso no estacionario, la falta de homogeneidad en los supuestos de distribución de cada observación hace necesaria alguna modificación para poder realizar procedimientos de comprobación de modelos cuando se supone que los datos tienen una distribución idéntica. Por lo general, sólo es posible aplicar estas comprobaciones de diagnóstico a una versión normalizada de los datos, condicionada a los valores de los parámetros ajustados. Por ejemplo, a partir de un modelo estimado

$$Z_t \sim \text{GEV}(\hat{\mu}(t), \hat{\sigma}(t), \hat{\xi}(t)),$$

las variables estandarizadas  $\tilde{Z}_t$ , definidas por

$$\tilde{Z}_t = \frac{1}{\hat{\xi}(t)} \log \left\{ 1 + \hat{\xi}(t) \left( \frac{Z_t - \hat{\mu}(t)}{\hat{\sigma}(t)} \right) \right\},$$

tienen cada una la distribución estándar de Gumbel, con función de distribución de probabilidad

$$\Pr\{\tilde{Z}_t \leq z\} = \exp\{-e^{-z}\}, \quad z \in \mathbb{R}. \quad (7)$$

Esto significa que se pueden hacer gráficos de probabilidad y de cuantiles de los  $\tilde{z}_t$  observados con referencia a la distribución (7). Denotando los valores ordenados de  $\tilde{z}_t$  por  $\tilde{z}_{(1)}, \dots, \tilde{z}_{(m)}$ , el diagrama de probabilidad está formado por los pares

$$\{i/(m+1), \exp(-\exp(-\tilde{z}_{(i)}))\}; \quad i = 1, \dots, m\},$$

mientras que el gráfico de cuantiles se compone de los pares

$$\{(\tilde{z}_{(i)}, -\log(-\log(i/(m+1))))\}; \quad i = 1, \dots, m\}.$$

El diagrama de probabilidad es invariable con respecto a la elección de Gumbel como distribución de referencia, pero el diagrama de cuantiles no lo es: otras elecciones distintas de Gumbel darían lugar a un diagrama diferente. A pesar de esta arbitrariedad, la elección de Gumbel es posiblemente la más natural dado su estatus dentro de la familia GEV.

Se pueden adoptar técnicas similares para la distribución de Pareto generalizada. En este caso, tenemos un conjunto de umbrales  $u(t)$  que posiblemente varíen en el tiempo, lo que da lugar a excesos de umbral  $y_{t_1}, \dots, y_{t_k}$ .<sup>II</sup> El modelo estimado, en su forma general, es

$$Y_t \sim \text{GP}(\hat{\sigma}(t), \hat{\xi}(t)).$$

Esta vez, dado que la distribución exponencial es un caso especial de la familia de Pareto generalizada con  $\xi \rightarrow 0$ , es más natural aplicar una transformación a una distribución exponencial estándar:

$$\tilde{Y}_{t_k} = \frac{1}{\hat{\xi}(t)} \log \left\{ 1 + \hat{\xi}(t) \left( \frac{Y_{t_k} - u_t}{\hat{\sigma}(t)} \right) \right\}.$$

Denotando los valores ordenados de los  $\tilde{Y}_{t_j}$  observados por  $\tilde{y}_{(1)}, \dots, \tilde{y}_{(k)}$  se deduce que se puede formar un diagrama de probabilidad con los pares

$$\{(i/(k+1), 1 - \exp(-\tilde{y}_{(i)}))\}; \quad i = 1, \dots, k\},$$

y un gráfico de cuantiles por los pares

$$\{(\tilde{y}_{(i)}, -\log(1 - i/(k+1)))\}; \quad i = 1, \dots, k\}.$$

Al igual que con la distribución GEV, el gráfico de probabilidad es invariable con respecto a la elección de la distribución de referencia, pero el gráfico de cuantiles es específico de la elección de la escala exponencial.

<sup>II</sup>La notación  $y_{t_j}$  se utiliza aquí para enfatizar el hecho de que es poco probable que el  $j$ -ésimo exceso de umbral corresponda a la  $j$ -ésima observación en el proceso original.

## Capítulo 3

# Metodología propuesta para el análisis de la evolución de la velocidad del viento

En este capítulo se presentan los recursos y procedimientos necesarios para poder estimar un modelo fiable que nos permita abordar el objetivo dado en la introducción. Para ello, se presenta una estrategia de modelización acorde con los resultados obtenidos en el análisis exploratorio.

### 3.1. Análisis exploratorio

Inicialmente se realiza un trabajo de análisis exploratorio sobre nuestro conjunto de datos asociado al observatorio del Aeropuerto de Reus, con indicador X0016A, haciendo una descriptiva de la velocidad, distinguiendo por meses. Para ello, se requieren los paquetes **kableExtra** para las visualizaciones de tablas, y **dplyr** junto con la función **group\_by** para agrupar en meses. Para especificar las funciones de resumen requeridas para las tablas, se recurre a **summarize**. Esta descriptiva está compuesta por la media, desviación típica, mediana, y coeficientes de variación y asimetría.

Seguidamente, se representan las gráficas de la descriptiva, mediante un **ggplot**, ajustando las curvas con **geom\_smooth** usando el método **loess** con bandas de confianza del 95 %. Todas las gráficas de este trabajo se han ajustado de esta manera. Se muestran también gráficos de los cuantiles 25, 50, 75 y 90 asociados al primer decenio (correspondiente desde 1961 hasta 1970) y al último decenio (correspondiente desde 2010 hasta 2019) de nuestro conjunto de datos, junto con su diferencia. Se termina este análisis de Reus con un **histograma**, que se realiza con los datos originales, con **ggplot** y la orden **geom\_histogram**, filtrando según sea Enero o Julio, y un **diagrama de cajas** con **ggplot** y **geom\_boxplot**. Además, para hacer algunas gráficas por grupos tanto en este apartado como en el resto, se utilizan los paquetes **reshape2** y **forcats** para poner las diferentes variables en una columna (con la orden **melt**) y reordenar columnas entre otros procedimientos para luego dibujar por grupos.

Finalmente, se realizan las gráficas de la media y desviación típica en las estaciones de las costas Nororiental y Este, para hallar posibles patrones estacionales regionales. Para hacer el análisis, se siguen los mismos pasos que el análisis que se realiza en Reus, creando las tablas de datos a nuestra conveniencia, y posteriormente dibujando la media y la desviación típica de la velocidad máxima de forma suavizada. Para ello se crean las tablas de cada zona, juntando los datos de cada zona en un nuevo conjunto de datos para poder hacer las gráficas por grupos de la misma forma que las gráficas anteriores.

### 3.2. Modelización inicial

En este apartado se analizan modelos iniciales que nos den pautas para poder encontrar un modelo óptimo que nos permita conocer mejor cómo serán las tendencias. Esta búsqueda de modelo, se realiza para Reus, y para Almería, con otro carácter meteorológico.

Para ambas estaciones, se consideran dos modelos `fevd`, uno con distribución `Gumbel`, y otro `GEV`, para compararlos y ver cuál obtiene mejor grado de ajuste. Para ello, previamente se cargan los paquetes `kableExtra`, `Lmoments`, `carData`, `car`, `distillery`, `extRemes`.

Se aplica la función `fevd` a nuestra columna `X0016A` para obtener los dos modelos para cada mes, y se realiza un bucle con `for` para almacenar en una tabla los valores  $AIC^I$  de ambos modelos (menores valores de  $AIC$  implica mejores ajustes), como los parámetros estimados de localización  $\mu$ , escala  $\sigma$  y forma  $\xi$  según el modelo y los intervalos de confianza asociados al valor del parámetro de forma en el modelo `GEV` sin distribución `Gumbel`.

Seguidamente se realiza una tabla donde se podrá ver si los parámetros de escala y localización dependen o no del año. Para ello, para cada mes, se tiene una columna con el valor de  $\mu_1$  del modelo `GEV` con `location.fun = ~ year`, el p-valor del test entre este modelo (con tendencia  $\mu_1$ ) y el modelo inicial `GEV` ( $\mu_1 = 0$ ), otras cuatro columnas con los niveles de retorno (NDR) a 10 y 50 años en 1961 y en 2019, es decir, el primer y último año de los datos que se disponen, y dos columnas con el cambio entre los niveles de retorno entre ambos años. Estos niveles de retorno se obtienen en todo este trabajo con `return.level`.

Finalmente, para ver de una forma más intuitiva cómo se comporta el coeficiente  $\mu_1$  en el resto de estaciones, se procede a dibujar en una gráfica los valores de  $\mu_1$  en cada mes para cada una de las estaciones, tanto de la Costa Nororiental como de la Costa Este.

Para ello, primero se realizan las tablas (mostradas en el Anexo A.E.4) que contienen los valores de  $\mu_1$  a lo largo del año para las Costas Nororiental y Este, de forma análoga a las tablas anteriores. Para hacer las gráficas, se instalan los paquetes `reshape2` y `forcats`, para poner las diferentes variables en una columna (con la orden `melt`) y reordenar columnas entre otros procedimientos, para luego dibujar por grupos mediante un `ggplot` con el método `loess`, mostrando los puntos significativos con `geom_point` y `geom_text`.

### 3.3. Modelización óptima

Se busca el mejor modelo para tres estaciones, la del Aeropuerto de Reus, una estación con notable tendencia, la del Aeropuerto de Almería, al mostrar menos tendencia, y la de Barcelona Fabra por su notoriedad.

Para cada estación, se empieza considerando un modelo inicial. El objetivo es obtener el modelo mejor ajustado, luego se recurre a los valores  $AIC$  y  $BIC$ . Se realizan tres tablas con los valores  $AIC$  y  $BIC$  de los modelos considerando diferentes combinaciones para los parámetros de localización ( $\mu$ ), escala ( $\sigma$ ) y forma ( $\xi$ ), con y sin interacción para considerar estacionalidad. La primera tabla es para los modelos en los que solamente interviene un parámetro de los tres, la segunda para los que intervienen dos, y la tercera para los que intervienen todos. Estas tres tablas se pueden ver en el Anexo A.F.4.

Seguidamente se realiza otra tabla en la que aparezcan los modelos con menor valor  $AIC$  de cada una de las tablas anteriores, junto con su valor  $AIC$  y  $BIC$  correspondiente, para obtener el modelo con menor  $AIC$ , es decir, el mejor ajustado. Se compara con el modelo inicial. En caso de mejorar se elige el modelo final, y si empeora se busca una explicación para considerar otro modelo.

---

<sup>I</sup>Los valores  $AIC$  (Criterio de Información de Akaike) y  $BIC$  (Criterio de Información Bayesiano) miden la bondad de ajuste de un modelo estadístico.



Tanto en los modelos iniciales como en los finales, además de ver su valor AIC, también se ven los niveles de retorno estimados para 1961 y 2019 a 2, 10 y 50 años, en fechas del año, gráficos de diagnóstico estándar, así como de los valores de los parámetros que intervienen en el modelo final a lo largo del año, junto con sus pendientes. También se compara la gráfica de los cuantiles empíricos tanto en los modelos inicial y final, junto con una vertical con el valor del cuantil empírico asociado a un valor elegido, cercano al punto donde se aprecia que produce un desvío.

### 3.4. **extRemes 2.0: un paquete de análisis de valor extremo**

Este apartado describe el paquete R **extRemes** versión 2.0 [8], con el que se van a ajustar los modelos GEV. Las funciones proporcionan utilidades para estimar modelos EVA, con un enfoque en aplicaciones meteorológicas y climáticas, incluida la incorporación de covariables.

El paquete **ismev** fue desarrollado para completar el libro de Coles (2001) [1] y se adopta su notación. Debido a la importancia de Coles (2001) se decidió mantener **extRemes** lo más cerca posible del original para que pueda continuar usándose junto con este texto.

Las herramientas de **extRemes** incluyen el tratamiento de los extremos univariados, tutoriales con el uso de datos meteorológicos. Además, en el contexto del software libre R [6], un gran número de usuarios proporcionan retroalimentación para la mejora continua.

Las herramientas que se van a utilizar en el ajuste de modelos son las siguientes:

- La función `fevd`, usa la variable respuesta para ajustar una distribución GEV. Incluye la distribución `Gumbel`, con la opción `type = "Gumbel"`. Permite usar el esquema usual de la función `lm` para definir la estructura del predictor lineal si se quiere dependencia en alguno de los parámetros de localización, escala y forma, con las opciones `location.fun`, `scale.fun` o `shape.fun`.
- La función `return.level` permite mostrar los niveles de retorno. Se obtienen los intervalos de confianza con la opción `do.ci = TRUE`, o se usa la función `ci` con `return.period` junto con los valores deseados.
- La función `lr.test`, realiza un test de cociente de verosimilitudes para comparar dos objetos `"fevd"`.
- Para calcular estimaciones del nivel de retorno se usan las funciones `make.qcov` con los valores que correspondan en cada caso, junto con `return.level` y `return.period` con los valores que nos interesen.



# Capítulo 4

## Resultados

### 4.1. Análisis exploratorio de los datos

Se realiza un análisis exploratorio sobre nuestro conjunto de datos asociado al observatorio del Aeropuerto de Reus, con indicador X0016A, haciendo una descriptiva de la velocidad, distinguiendo por meses, junto con gráficos de los cuantiles 25, 50, 75 y 90. Además, se resume la evolución en las regiones mediterráneas correspondientes a las Costas Este y Nororiental. El resultado del resto de observatorios se pueden encontrar en el Anexo A.D.2.

#### 4.1.1. Aeropuerto de Reus

La distribución empírica se resume por meses en la tabla 2 ubicada en el Anexo A.D.1, haciendo una descriptiva de la velocidad máxima diaria, con media, desviación típica, mediana y coeficientes de variación y asimetría, mostradas en la figura 4.1 y la figura 4 del Anexo A.D.1. Destaca el coeficiente de asimetría positivo, 0.44 en Junio (mínimo), y 0.90 en Mayo (máximo), siendo una distribución fuertemente asimétrica.

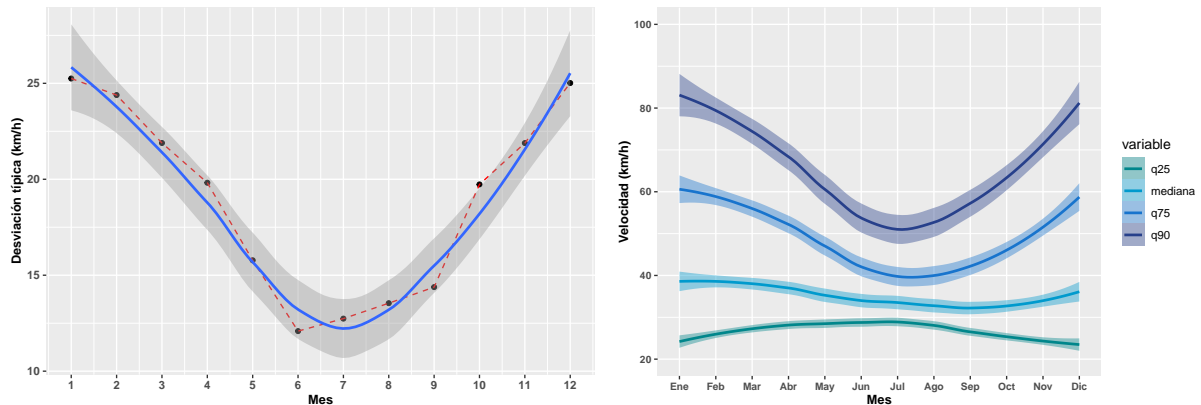


Figura 4.1: Desviación típica (izda.) y cuantiles 25, 50, 75 y 90 (dcha.) de la velocidad máxima expresada en km/h para cada mes en Reus. En ambos casos las curvas se han ajustado con `geom_smooth` usando el método `loess` con bandas de confianza del 95 %.

La desviación típica en este estudio va a ser crucial, ya que muestra la variación de los datos según en qué mes estemos, y en qué magnitud. En Enero y en Diciembre, los datos pueden ser muy variables, con valores muy altos de velocidad máxima y días con ausencia de viento, con un valor de desviación típica de aproximadamente 25. En verano hay menos variabilidad, ya que el valor obtenido de desviación típica es cercano a 12. Esta variabilidad indica una marcada estacionalidad.

Para la gráfica de cuantiles, podemos ver que las mayores diferencias se producen en invierno, con situación opuesta en verano. En verano, el rango intercuartílico se estrecha porque en Julio q75 alcanza su mínimo y también q25 su máximo, es decir, en Enero los vientos débiles son más frecuentes que en Julio.

La figura 6 del Anexo A.D.1 muestra los **coeficientes de variación y asimetría** en cada mes. Por un lado, la variable es definida positiva, es conveniente usar como medida de dispersión relativa el coeficiente de variación. Este coeficiente es mayor en Diciembre y Enero, menor en Julio. Por otro lado, la distribución es fuertemente asimétrica, menos en Julio y más en Marzo y Noviembre. Los inviernos suelen tener más movimiento meteorológico, mientras que los veranos suelen ser más calmados.

Las figuras 7 y 8 del Anexo A.D.1 muestran un **histograma** de datos para Enero y Julio, que son los meses más extremos, un **diagrama de cajas** para todos los meses, y las gráficas frente a los días del año.

### Evolución temporal

La figura 4.2 muestra cuantiles 25, 50, 75 y 90 por meses, para el primer y último decenio. Se incluye un **loess** para ayudar en la interpretación. Las gráficas de la figura 5 del Anexo A.D.1 muestran la diferencia de los valores entre ambos periodos para cada cuantil. Se encuentran evidencias del decaimiento, en consonancia con los resultados del suplemento especial del Boletín de la Sociedad Meteorológica Americana, "*State of the Climate in 2022*" [5]. Hay disminuciones notorias en los percentiles más elevados.

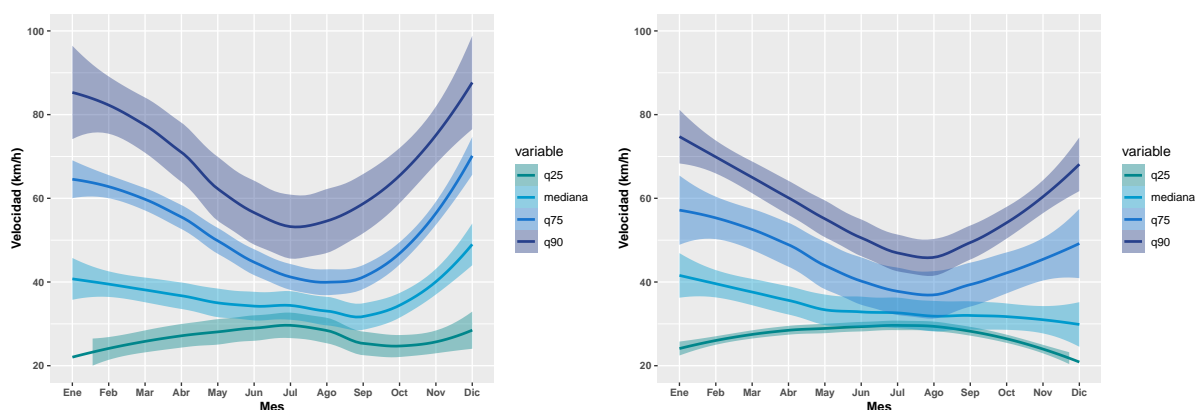


Figura 4.2: Cuantiles 25, 50, 75 y 90 de la velocidad máxima para cada mes en Reus durante los periodos 1961 – 1970 (izda.), 2010 – 2019 (dcha.). En ambos casos las curvas se han ajustado con `geom_smooth` usando el método `loess` con bandas de confianza del 95 %.

Para q25, los cambios son cercanos al 0 todo el año, salvo en los últimos meses del año. Para q50, los valores son negativos cercanos a 0 todo el año, descendiendo acusadamente a partir de Septiembre. Para q75, los valores son negativos siendo cercanos a 0 en Septiembre, y son más acusados a partir de este mes. Para q90, los cambios son bastante negativos durante todo el año, siendo más acusados en verano, y menores en los meses finales del año. Estos hechos inducen a que hay un cambio en la distribución, que en la mediana es negativo y próximo a 0 y en el cuantil 90 es más marcado, por ello el modelo con distribución Gumbel puede ser demasiado rígido.

#### 4.1.2. Costas Nororiental y Este

Se realiza el análisis exploratorio de dos regiones distintas en las que se ven diferentes patrones estacionales, pese a estar próximas geográficamente.

⊗ **COSTA NORORIENTAL**  $\Rightarrow$  Cordillera costera catalana. Tortosa (X9981A), Aeropuerto de Reus (X0016A), Aeropuerto de Barcelona (X76) y Barcelona Fabra (X0200E).

⊗ **COSTA ESTE**  $\Rightarrow$  Cordillera Ibérica al oeste, llanuras costeras. Aeropuerto de Valencia (X8414A), Valencia Viveros (X8416) y Castellón - Almassora (X8500A).

La figura 4.3 muestra la velocidad máxima media y la desviación típica según el día del año de cada estación (las 9 restantes se encuentran en el Anexo A.D.2). Las estaciones con mayores vientos son las de Tortosa y Reus, por su localización en el valle del Ebro y el cierzo asociado a éste, caracterización que ocasiona la forma de V más pronunciada en la gráfica de la desviación típica. Hay valores menores en el resto de estaciones, sobre todo en verano. Se observa una marcada estacionalidad, siendo los valores medios mayores en la Costa Nororiental, superando ampliamente a Castellón y Viveros. Los mínimos de ambas zonas se producen en verano, siendo los máximos en Enero en la Costa Nororiental, y en Marzo en la Costa Este. Esto nos indica un patrón de **estacionalidad distinta por áreas**, es decir, el mismo modelo no nos va a dar la misma estacionalidad según el lugar en el que nos encontremos.

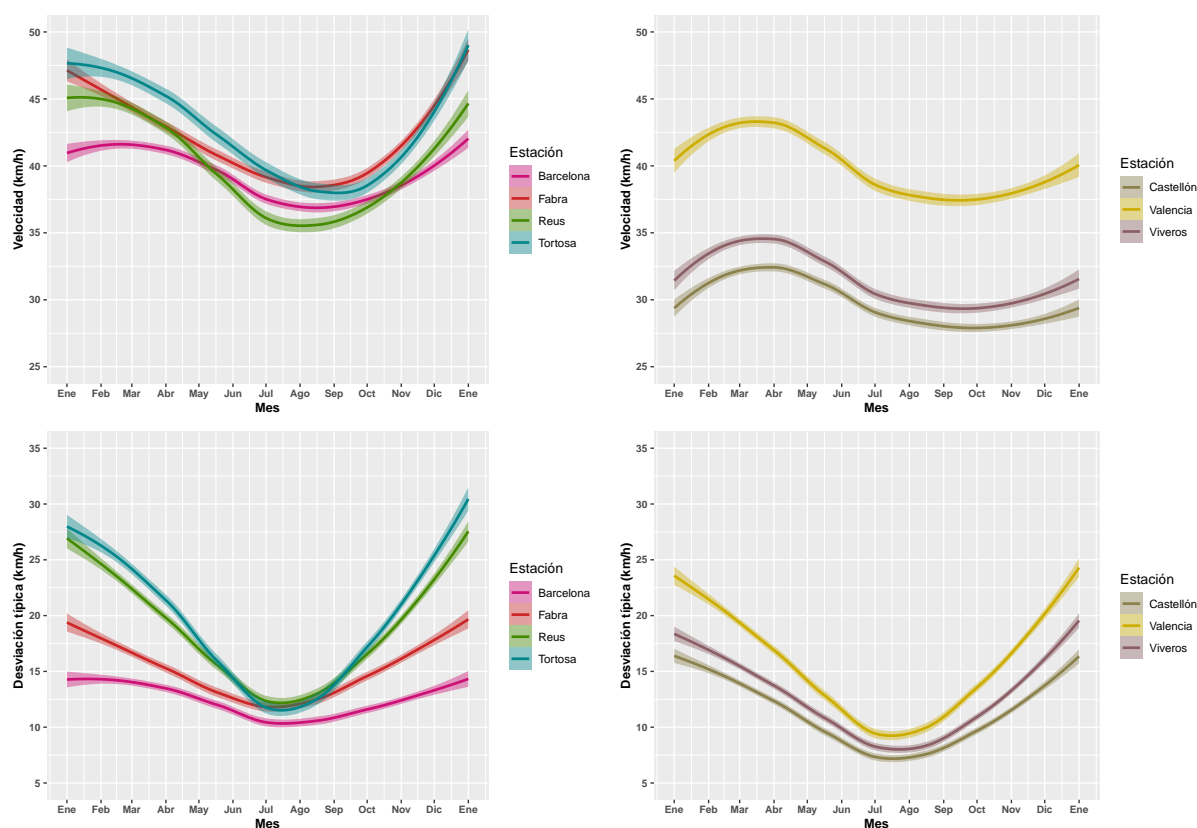


Figura 4.3: Velocidad máxima media (arriba) y desviación típica (abajo) según el día del año para las estaciones de la Costa Nororiental (izda.) y Costa Este (dcha.). Se muestra la curva loess y sus bandas de confianza del 95 %.

## 4.2. Modelización inicial

Se ha visto una marcada estacionalidad en Reus, influenciada por el viento provocado por el delta del Ebro. Este hecho provoca que la variabilidad de los datos sean dispares según la época del año, luego es una estación con bastante movimiento meteorológico.

Todos estos aspectos van a ser necesarios para poder encontrar un modelo óptimo para estimar las tendencias. Esta búsqueda de modelo, vamos a hacerla para Reus, y para Almería

(detallada en el Anexo A.E.2), con menor tendencia. Para cada estación, se consideran dos modelos `fevd`, uno con distribución `Gumbel` (denotado G), y otro GEV (denotado S), para comparar los valores obtenidos entre ellos y ver cuál es más recomendable para trabajar. El modelo estimado para las regiones definidas por las Costas Nororiental y Este se muestra en el Anexo A.E.3.

#### 4.2.1. Aeropuerto de Reus

La tabla 4.1 (tabla 3 completa para todos los meses del año en el Anexo A.E.1) nos muestra que hay una ganancia en el modelo sin distribución `Gumbel`, y además el mes de Agosto es el único en el que  $\xi$  puede ser compatible con el valor 0. Además,  $\xi$  - S muestra variaciones también, siendo más altos en meses de invierno, y más bajos en los meses de verano, hecho que implica estacionalidad, y por tanto los tres coeficientes muestran estacionalidad. Para el intervalo de confianza de  $\xi$ , vemos que solamente contiene al 0 el mes de Agosto.

Mes	AIC - S	AIC - G	$\mu$ - S	$\mu$ - G	$\sigma$ - S	$\sigma$ - G	$\xi$ - S	IC $\xi <$	IC $\xi >$
Enero	16519.81	16581.14	31.35	33.17	16.85	18.52	0.19	0.14	0.24
Febrero	14917.70	14962.08	32.80	34.21	16.55	17.78	0.15	0.11	0.20
Julio	14129.99	14137.40	30.93	30.74	10.15	10.13	-0.04	-0.06	-0.01
<b>Agosto</b>	14258.55	14256.56	30.32	30.33	10.22	10.23	<b>0.00</b>	<b>-0.02</b>	<b>0.03</b>

Tabla 4.1: Valores AIC de ambos modelos, valores de los parámetros de localización  $\mu$ , escala  $\sigma$  y forma  $\xi$  según el modelo y los intervalos de confianza asociados al valor del parámetro de forma en el modelo sin distribución `Gumbel` (S). Los intervalos de confianza de  $\xi$  que contienen al 0 están marcados en negrita junto con el valor de  $\xi$  y el mes correspondiente.

A la vista de los resultados, los modelos indican que es mejor el modelo sin distribución `Gumbel`, ya que los modelos arrojan resultados similares pero el de distribución `Gumbel` que tiene el parámetro de forma  $\xi = 0$  es más rígido.

La tabla 4.2 mostrará resultados para valorar que esta modelización es insuficiente, y que por lo tanto deberán plantearse modelos en los que los parámetros de escala y localización dependan del año. La tabla muestra resultados para 4 meses, y se encuentra completa en la tabla 4 del Anexo A.E.1. Para cada mes, aparece una columna con la tendencia estimada,  $\mu_1$ , del modelo GEV con `location.fun = ~ year`, el p-valor asociado a la hipótesis nula de modelo estacionario GEV ( $\mu_1 = 0$ ); las otras columnas muestran los niveles de retorno (NDR) a 10 y 50 años en 1961 y en 2019 y el cambio entre los niveles de retorno entre ambos años.

Para comprender la tabla, notar que si el valor de  $\mu_1 = 0$ , entonces no habría cambio temporal, si  $\mu_1 < 0$  entonces disminuiría el efecto viento y si  $\mu_1 > 0$  aumentaría. Por otro lado, para el test de hipótesis, la hipótesis nula consiste en tener un parámetro de localización  $\mu_0$  constante para todos los años frente a una alternativa de que  $\mu$  depende del tiempo y va a ser un cierto  $\mu_0 + t\mu_1$ , siendo el valor  $\mu_1$  la pendiente que tenemos estimada. Si el p-valor del test no rechaza  $\mu_1 = 0$ , es decir, p-valor  $> 0,05$ , entonces no hay tendencia; de lo contrario, p-valor cercano a 0 indica que sí habría tendencia, con cambios en los niveles de retorno entre 1961 y 2019. A la vista de los cambios y según sea el signo de  $\mu_1$  sabremos si aumenta o disminuye.

En la tabla observamos que  $\mu_1 > 0$  en todos los meses salvo en Diciembre, y tenemos un p-valor cercano a 0 salvo en Enero y Diciembre. Además los niveles de retorno muestran que éstos serían mayores en 2019, en prácticamente todo el año hay tendencia de que haya un aumento en la velocidad del viento, con cambios de entre 1 y 7 km/h, siendo este aumento más acusado en los meses de verano. Se observa que los cambios para NDR a 10 y 50 coinciden, vamos a ver porqué:

En el Capítulo 2 hemos visto que  $z_p$  es el **nivel de retorno** asociado al **periodo de retorno**  $1/p$ , se espera que el nivel  $z_p$  se supere por término medio una vez cada  $1/p$  años. Más

Mes	$\mu_1$	p-valor	NDR-10		Cambio	NDR-50		Cambio
			1961	2019	10	1961	2019	50
Enero	0.04	1.00	77.85	80.37	2.53	130.73	133.26	2.53
Febrero	0.03	0.00	75.68	77.58	1.90	120.28	122.18	1.90
Julio	0.09	0.00	50.01	55.44	5.43	65.93	71.36	5.43
Agosto	0.07	0.00	51.20	55.51	4.31	69.03	73.34	4.31

Tabla 4.2: Valor de  $\mu_1$  del modelo GEV con `location.fun = ~ year`, p-valor del test entre este modelo (con tendencia  $\mu_1$ ) y el modelo inicial GEV ( $\mu_1 = 0$ ), niveles de retorno (NDR) a 10 y 50 años en 1961 y en 2019, es decir, el primer y último año de los datos que disponemos respectivamente y cambio entre los niveles de retorno entre ambos años.

concretamente,  $z_p$  es superado por el máximo anual con probabilidad  $p$ .

En particular, definiendo  $y_p = -\log(1 - p)$ , y teniendo ahora en cuenta el año elegido como  $t$  obtenemos que

$$z_p^t = \begin{cases} \mu^t - \frac{\sigma}{\xi}[1 - y_p^{-\xi}], & \text{para } \xi \neq 0, \\ \mu^t - \sigma \log y_p, & \text{para } \xi = 0, \end{cases}$$

donde  $\mu^t = m_0 + m_1 \cdot t$ .

Por tanto, para dos años  $t_1$  y  $t_2$ , si  $\xi \neq 0$ , tenemos que:

$$z_p^{t_2} - z_p^{t_1} = \mu^{t_2} - \frac{\sigma}{\xi}[1 - y_p^{-\xi}] - \mu^{t_1} + \frac{\sigma}{\xi}[1 - y_p^{-\xi}] = m_0 + m_1 \cdot t_2 - m_0 - m_1 \cdot t_1 = m_1 \cdot (t_2 - t_1).$$

Si  $\xi = 0$  ocurre lo mismo:

$$z_p^{t_2} - z_p^{t_1} = \mu^{t_2} - \sigma \log y_p - \mu^{t_1} + \sigma \log y_p = m_0 + m_1 \cdot t_2 - m_0 - m_1 \cdot t_1 = m_1 \cdot (t_2 - t_1).$$

Esto nos muestra que el cambio es el mismo para cualquier día, ya que nuestro modelo lo impone. Esto hace demasiado rígido este modelo inicial.

### 4.3. Modelización óptima

Se pretende ajustar el mejor modelo para tres estaciones, Aeropuerto de Reus (bastante tendencia), Aeropuerto de Almería (poca tendencia), y Barcelona Fabra.

Para ello, para cada estación, empezaremos considerando un modelo inicial. Como hemos visto que hay estacionalidad y tendencia en los valores de los datos de las 3 estaciones y en sus tres coeficientes, veremos un modelo en el que la respuesta dependa del año. Luego en `location.fun` tendremos que poner una expresión que sea periódica anual, para ello consideramos el seno y el coseno del primer armónico de la expresión siguiente, donde  $t$  depende del día del año, y  $T$  es el número total de días del año, 366, ya que puede haber años bisiestos:

$$c_1 = \cos \frac{2\pi}{T}t \quad s_1 = \sin \frac{2\pi}{T}t$$

Además, hay que tener en cuenta que como tenemos tendencia y estacionalidad, el efecto de la variable `year` tiene que ser cambiante a lo largo de los meses. Luego consideramos `year + c1 + s1`, donde la pendiente es idéntica para cada mes, y un segundo modelo en el que la pendiente cambie para cada fecha del año, con interacción entre tendencia y estacionalidad `location.fun = year*(c1 + s1)`.

### 4.3.1. Aeropuerto de Reus

Comenzamos considerando el **modelo inicial**, estimado con `fevd` en el que intervienen los datos del Aeropuerto de Reus, y `location.fun = ~ year + c1 + s1`, es decir, sólo interviene el parámetro de localización sin interacción, con unidades en km/h. De este modelo obtenemos un **valor AIC** de 182183,2.

Ahora vamos a ver los **niveles de retorno** estimados para 1961 y 2019 a 2, 10 y 50 años. El de 2 es la mediana, la mitad de valores por encima y por debajo. Aunque la tendencia es única, el coseno y el seno de los armónicos influye en la localización, luego en invierno y en verano sus curvas estarán en diferentes posiciones.

La tabla 9 del Anexo A.F.1 incluye los niveles de retorno estimados. Los valores de retorno más elevados se obtienen en primavera y los menores en invierno. Las rigideces de este modelo inicial lo hacen no útil, porque no incluye interacción entre tendencia y estacionalidad y por lo tanto estima el mismo cambio a lo largo del año, es decir, solo incluye tendencia en la localización, y da lugar al mismo cambio para todos los valores de retorno.

La tabla 4.3 contiene los valores AIC más bajos obtenidos en cada una de las tablas del Anexo A.F.4, comparando distintas modelizaciones. El objetivo es identificar el modelo final.

Valor	$2\sigma$	$2\mu + 1\sigma$	$2\mu + 2\sigma + 2\xi$
AIC	179821.2 (+0.33 %)	179499.2 (+0.15 %)	<b>179218.9</b> (100 %)
BIC	179885.1 (+0.29 %)	179587.0 (+0.12 %)	<b>179362.5</b> (100 %)

Tabla 4.3: Valores AIC más bajos obtenidos en cada una de las tablas de modelos para el Aeropuerto de Reus del Anexo A.F.4.

El menor AIC (179218,9) corresponde al modelo con `location.fun = ~ year*(c1 + s1)`, `scale.fun = ~ year*(c1 + s1)` y `shape.fun = ~ year*(c1 + s1)`, es decir, modelo en el que los tres parámetros, localización  $\mu$ , escala  $\sigma$  y forma  $\xi$ , tienen expresión `year*(c1 + s1)`.

El parámetro de forma va a tener un comportamiento estacional y va a cambiar con el tiempo. Este va a ser nuestro modelo finalista, en el que viendo en R-studio [7] un resumen obtenemos valores de  $\mu_1 = -2,562225 \cdot e^{-02}$ ,  $\sigma_1 = -1,556943 \cdot e^{-01}$  y  $\xi_1 = 3,228816 \cdot e^{-03}$ .

Para ver la tendencia general, vemos que el valor  $\mu_1$  es negativo, además  $\sigma_1$  también baja y cada vez hay menos variabilidad, pero se está modificando la forma, fijándonos en el valor de  $\xi_1$ . Esta tendencia general además es compleja al tener interacción con seno y coseno.

Las figura 13 del Anexo A.F.1 muestra los gráficos de diagnóstico estándar de los modelos inicial y finalista para compararlos. De las gráficas anteriores, vamos a extraer la gráfica de los cuantiles empíricos para ver si mejora con respecto al modelo inicial. Para ello, se fija cuál de las dos tiene más valores por encima de 105 km/h, que es donde se desvían empíricos y estimados, y se considera el valor cuantil empírico asociado a ese valor, dibujando la vertical de ese valor en ambas gráficas. Cuanto más a la derecha esté esa vertical, mejor será el modelo.

Comparando ambas gráficas en la figura 4.4, aparece un comportamiento más satisfactorio en el de la derecha, el modelo que tiene menor AIC, que tiene interacciones. Por tanto, consideramos este último como el **modelo final**.

Los **niveles de retorno** estimados para 1961 y 2019 a 2, 10 y 50 años en algunas fechas permiten valorar el cambio en la distribución.

En la tabla 4.4 tenemos unos valores de retorno que son más elevados en torno a los primeros y últimos días del año, y menores en los días elegidos de pleno verano. En este modelo final, todos los cambios tienen tendencia a disminuir, siendo más acusados en los meses de valores de retorno más elevados, ocurriendo lo contrario para los menores.

En las tablas 4.2 (Sección 4.2.1) y 9 (Anexo A.F.1) observábamos que los cambios en los niveles de retorno aumentaban, usando un modelo muy restrictivo dado que su forma era rígida,



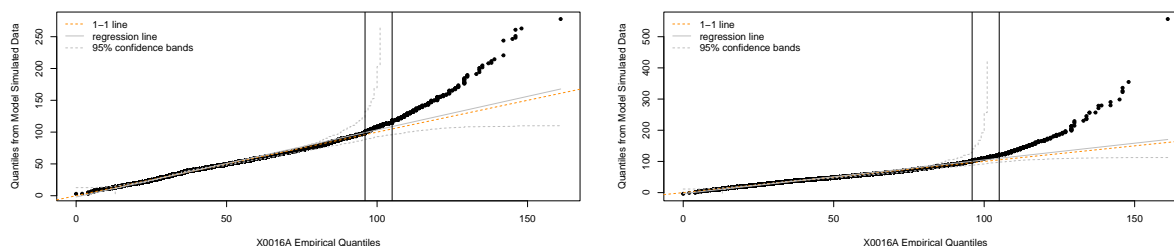


Figura 4.4: Cuantiles empíricos para los modelos inicial y final. Segmento vertical en  $c$  el valor 105, y en el cuantil empírico asociado a ese valor.

Fecha	NDR-2		Cmb	NDR-10		Cmb	NDR-50		Cmb
	1961	2019	2	1961	2019	10	1961	2019	50
1 Ene	41.42	33.19	-8.23	89.34	66.99	-22.35	143.05	117.61	-25.43
5 Mar	43.18	34.46	-8.72	84.50	65.35	-19.15	123.62	111.09	-12.54
15 Abr	41.35	34.79	-6.57	74.09	57.28	-16.81	100.68	87.09	-13.59
24 May	38.22	34.54	-3.68	64.31	49.58	-14.72	83.48	67.09	-16.39
24 Jun	35.65	34.02	-1.63	58.80	45.48	-13.33	75.55	57.76	-17.79
15 Ago	33.17	32.90	-0.27	57.27	44.12	-13.15	76.63	55.75	-20.87
12 Oct	34.80	32.19	-2.61	68.76	51.12	-17.64	102.55	73.46	-29.09
8 Dic	39.48	32.68	-6.80	85.26	63.07	-22.19	136.88	106.17	-30.72

Tabla 4.4: Niveles de retorno a 2, 10 y 50 años para 1961 y 2019 junto con los cambios.

pero ahora, al haber considerado un modelo más complejo, en la tabla 4.4 obtenemos cambios de niveles de retorno más fiables.

Seguidamente, vamos a ver una representación de los efectos hallados con el modelo final. Para ello, vamos a ver una representación de cómo es la pendiente temporal, expresada en cambio por 10 años (multiplicando por 10 el valor de la pendiente en cada gráfica), a lo largo de un año diariamente expresado en meses para interpretar mejor la gráfica, para cada uno de los parámetros. Además, veremos el valor de cada parámetro en 1961 y 2019 para mostrar su evolución, para ello estimaremos el parámetro en cada día de ambos años.

En la figura 4.5, observamos que: (a) El parámetro de localización  $\mu$ , tiene pendiente negativa salvo en el periodo de verano, donde es positiva. Sus valores se han modificado apreciablemente, en 1961 los valores máximos suceden en primavera y los menores a principios de otoño, mientras que en 2019 los valores máximos suceden en verano y los menores en invierno. (b) El parámetro de escala  $\sigma$ , tiene pendiente negativa todo el año, con valores bastante lejanos al 0, y los valores de su expresión disminuyen en 2019 con respecto a 1961 (aproximadamente una disminución de 10km/h), pero con el mismo comportamiento anual. (c) El parámetro de forma  $\xi$ , tiene pendiente positiva a lo largo de todo el año con valores muy cercanos al 0, es decir, la forma no cambia en prácticamente todo el año. Los valores de su expresión son más elevados en 2019 que en 1961, pero tampoco es una diferencia muy llamativa, y tampoco varía mucho el comportamiento, siendo un poco más acentuado en 1961 y con valores mínimos a principios de verano, siendo a finales de éste en 2019.

#### 4.3.2. Otros observatorios

Se resumen brevemente los resultados del Aeropuerto de Almería (completos en el Anexo A.F.2) y los de Barcelona Fabra (completos en el Anexo A.F.3).

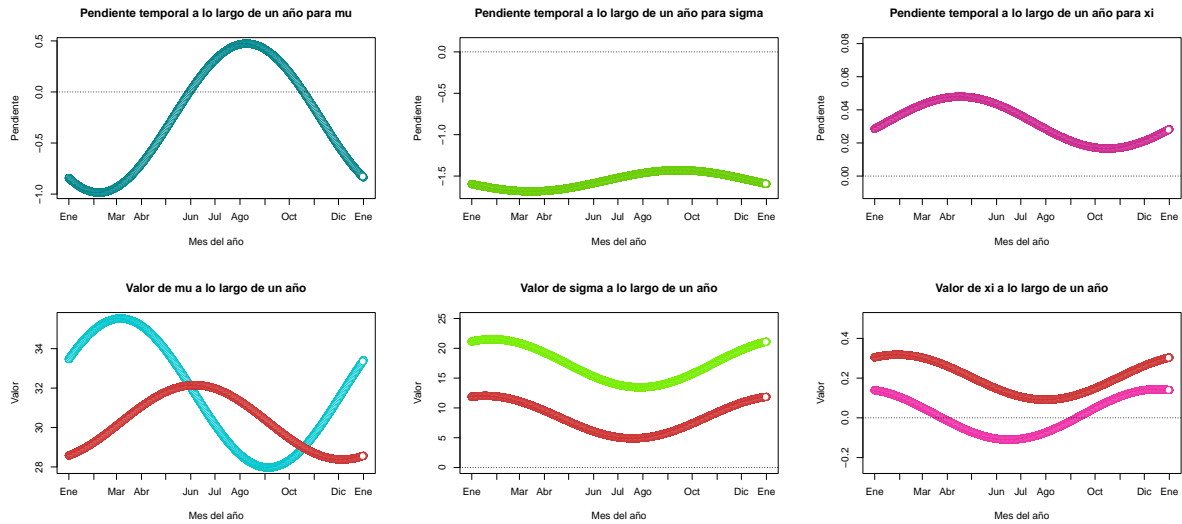


Figura 4.5: Arriba, cambio estimado por década de cada parámetro, la línea horizontal corresponde a 0. Abajo, parámetro estimado en cada día de 1961 (azul) y 2019 (rojo).

### Aeropuerto de Almería

El modelo obtenido con menor valor AIC de la tabla 11 del Anexo A.F.2 (tendencia en  $\mu$  y  $\xi$  con interacción, y  $\sigma$  sin interacción) muestra resultados peores en sus gráficos de diagnóstico estándar en la figura 14 del Anexo A.F.2 comparados con los del modelo inicial (tendencia en  $\mu$  sin interacción). Teniendo en cuenta el estudio realizado en la optimización inicial, estudiamos los otros dos modelos obtenidos en la tabla 11 del Anexo A.F.2 con el parámetro de forma  $\xi = 0$ , quedándonos finalmente con el modelo que tiene tendencia con interacción en el parámetro de escala  $\sigma$ .

Los valores de retorno obtenidos en la tabla 12 del Anexo A.F.2 tienen tendencia a disminuir, aumentando estos valores negativos para los niveles de retorno a una cantidad de años mayor, y siendo más acusados en los meses de invierno y menores en los de verano. En la figura 17 del Anexo A.F.2, observamos que el parámetro de escala  $\sigma$ , tiene pendiente negativa todo el año, y los valores de su expresión disminuyen en 2019 con respecto a 1961.

### Barcelona Fabra

El modelo obtenido con menor valor AIC de la tabla 14 del Anexo A.F.3, con tendencia en  $\mu$ ,  $\sigma$  y  $\xi$  con interacción, muestra mejores resultados en sus gráficos de diagnóstico estándar en la figura 18 del Anexo A.F.3, comparados con los del modelo inicial, con tendencia en  $\mu$  sin interacción.

En las figuras 20, 21 y 22 del Anexo A.F.3, observamos que: (a) El parámetro de localización  $\mu$ , tiene pendiente negativa salvo en invierno y finales de otoño, con lo que este parámetro tiene mayor variabilidad estacional en 2019. (b) El parámetro de escala  $\sigma$ , tiene pendiente negativa todo el año. (c) El parámetro de forma  $\xi$ , tiene pendiente negativa en prácticamente todo el año salvo en verano, con lo que su valor estimado es negativo para 2019, mientras que era positivo desde Noviembre a Marzo en 1961. Como resultado, los valores de retorno obtenidos en la tabla 15 del Anexo A.F.3 tienen tendencia a disminuir. A 2 años estas disminuciones son más acentuadas en verano y menores en invierno, ocurriendo lo contrario a 10 y 50 años.

## Capítulo 5

# Conclusiones

En este trabajo se han revisado los conceptos necesarios para plantear modelos asociados a valores extremos, en el ámbito de las distribuciones GEV no estacionarias. El objetivo es representar la distribución de las series de velocidad de viento máxima diaria, recogiendo sus características claramente estacionales, y particularmente estudiar la posible existencia de una disminución del viento, fenómeno conocido como “stilling” en el ámbito científico de cambio climático.

Se ha diseñado una metodología adecuada para abordar los objetivos. La propuesta incluye una fase exploratoria con estadísticos descriptivos y gráficas, una fase de modelización inicial donde hacemos uso de modelos de regresión GEV, con y sin distribución Gumbel, para decidir qué efectos pueden ser relevantes, estacionalidad, tendencia e interacciones, en los parámetros de localización, dispersión y forma. En la última fase se propone una modelización óptima, con modelos que nos permiten estudiar la evolución de la distribución, en particular de los valores de retorno.

Las herramientas exploratorias en este trabajo van a ser esenciales para deducir las características de la velocidad máxima diaria en observatorios peninsulares con diversos climas, con diferentes tendencias y regímenes. Se ha utilizado la librería de R ‘**extRemes 2.0**’ [8], que permite definir modelos para los parámetros de posición, forma y escala en distintas distribuciones. Concretamente, se han considerado variantes de modelos mostrados en el artículo *extRemes 2.0: An Extreme Value Analysis Package in R* [3].

### Conclusiones respecto al “stilling” en los observatorios del Mediterráneo

La modelización se ha aplicado a los observatorios de Aeropuerto de Reus, Aeropuerto de Almería y Barcelona Fabra, en el periodo 1961-2019. Se han estimado tendencias significativas en los parámetros de localización y forma en los 3 observatorios, indicando una concentración de la distribución en un rango más reducido al final del periodo. Aparecen tendencias significativas en los tres parámetros, que disminuyen en prácticamente todos los casos. Esta tendencia provoca incluso cambios en el carácter positivo o negativo del parámetro de forma. Los niveles de retorno a 2, 10 y 50 años en 1961 y 2019, estimados a partir de la modelización, tienden a disminuir en las tres estaciones meteorológicas, más acusadamente a 2 años en Reus, estación con más variabilidad, y a 50 años en Fabra, con disminuciones más marcadas que en el resto.

Para ayudar a la interpretación de estos resultados, se ejemplifica la evolución de los cuantiles en dos días, 15 de Abril y 15 de Agosto, comparando los valores estimados en 1961 y 2019, en las tres estaciones. En la tabla 5.1 se muestran los cuantiles 50 y 90. Todos los valores han disminuido entre ambos años, disminuciones más acentuadas en los valores altos de velocidades de viento, en general asociados a borrascas y tormentas.

Una visión complementaria se obtiene con el cambio de las funciones de densidad estimadas en ambos días, para ello usaremos la orden `devd` con los valores estimados de los parámetros para el 15 de Abril y el 15 de Agosto en ambos años. Obtenemos la figura 5.1, donde vemos

Cuantil	Aeropuerto de Reus		Aeropuerto de Almería		Barcelona Fabra	
	1961	2019	1961	2019	1961	2019
15 de Abril						
50	41.35	34.79	41.45	39.77	43.05	39.08
90	74.09	57.28	76.82	66.30	71.52	53.64
15 de Agosto						
50	33.17	32.90	40.44	38.86	39.45	34.37
90	57.27	44.12	70.52	60.62	63.06	47.06

Tabla 5.1: Cuantiles 50 y 90 en el Aeropuerto de Reus, Aeropuerto de Almería y Barcelona Fabra para el 15 de Abril y el 15 de Agosto de 1961 y 2019.

claramente el “stilling”, que se está notando en las velocidades altas. Además en las estaciones de los aeropuertos de Reus y Almería la distribución se ha desplazado a la izquierda. Esto nos indica que una evidencia de “stilling”, efecto del cambio global en el aspecto que nos ocupa.

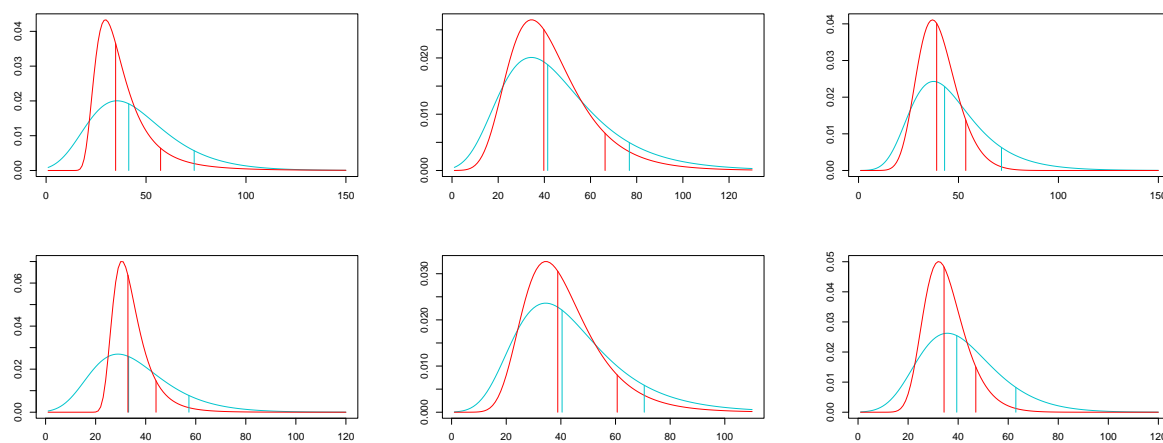


Figura 5.1: Funciones de densidad para el 15 de Abril (arriba) y el 15 de Agosto (abajo), Aeropuerto de Reus izda., Aeropuerto de Almería centro, Barcelona Fabra dcha. Los segmentos verticales representan q50 y q90. El color turquesa corresponde a 1961 y el rojo a 2019.

El cambio en la distribución no es igual a lo largo del año. La figura 5.2 muestra gráficas del cambio absoluto y cambio relativo en los cuantiles 50 (trazo rosa) y 90 (verde) para varios días. Notad que el cambio absoluto es la diferencia entre los cuantiles estimados en 1961 y 2019, y el cambio relativo es el ratio del cambio absoluto respecto al cuantil del año inicial. Los valores estimados se muestran en las tablas 25, 26 y 27 del Anexo A.G.1. Las gráficas, generadas por la función `ggplot`, incluyen un suavizado obtenido con el método `loess`.

En la figura 5.2 ha habido decrecimientos en todos los casos. En el cambio absoluto de q90 tanto en Reus como en Almería se aprecia un efecto estacional, que da lugar a mayores cambios en invierno (-30 km/h Reus, -15 km/h Almería) y menores en verano (-15 km/h Reus, -9 km/h Almería), por el contrario Barcelona Fabra tiene un cambio similar durante todo el año, próximo a 15 km/h. El cambio absoluto en q50 en Reus es muy apreciable y varía estacionalmente (cambios entre -10km/h y -1 km/h), los cambios en Barcelona y Almería son menos notorios (hasta -5 km/h) y tienen un comportamiento similar durante todo el año.

En estas condiciones, en términos absolutos, los cambios en la forma de la distribución más notables aparecen en invierno (Enero o Diciembre), y en el Aeropuerto de Reus, dado que el cambio en q50 y q90 se diferencia en 21 km/h en Enero. En términos relativos, el cambio en q50

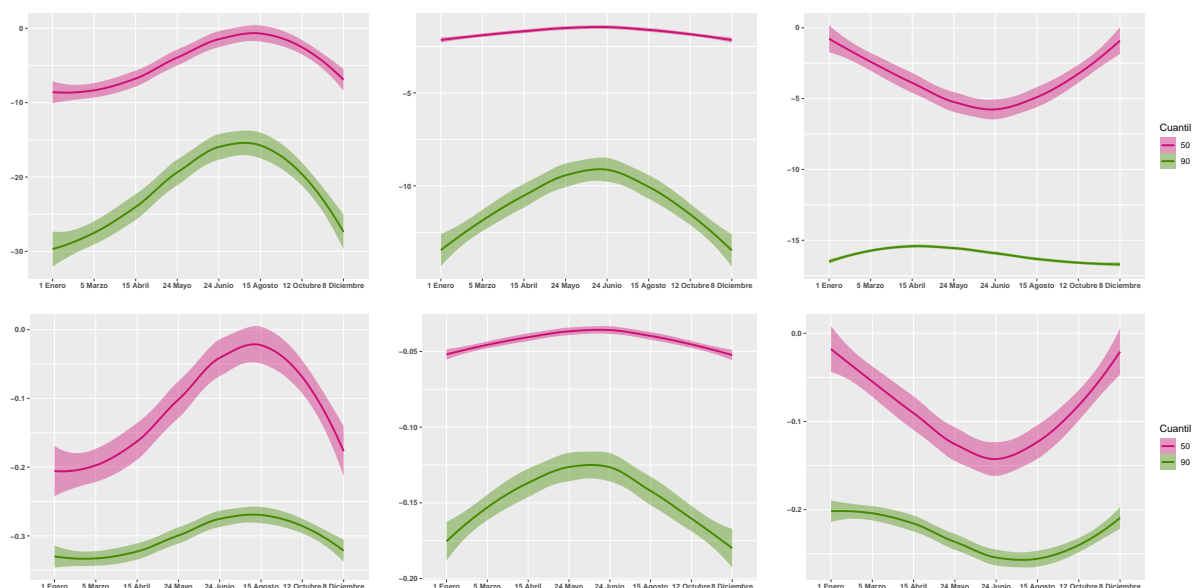


Figura 5.2: Cambio absoluto (arriba) y relativo (abajo) en el Aeropuerto de Reus izda., Aeropuerto de Almería centro y Barcelona Fabra dcha. El trazo rosa representa el cambio en  $q_{50}$  y el trazo verde en  $q_{90}$ .

es poco relevante en Almería durante todo el año, en verano en Reus y en invierno en Barcelona; en estos dos observatorios, el cambio relativo más importante es del orden de 15-20 %, en invierno en Reus y en verano en Barcelona. La reducción relativa en  $q_{90}$  para Reus está próxima al 30 % durante todo el año, pero es del entorno del 20 % en Barcelona y del 15 % en Almería.

En la figura 23 del Anexo A.G.2 se observan dos gráficas con los cambios absolutos y relativos en las tres estaciones.

### Líneas futuras

La metodología que se ha propuesto en este TFG ha permitido analizar el efecto de tendencias temporales diferentes para épocas del año que afectan no solo a un parámetro de posición, y de escala, sino que son capaces de estimar cambios en el parámetro de forma. Esto es especialmente importante para el análisis de variables con distribución fuertemente asimétrica, como es el caso analizado de la velocidad de viento máxima diaria. Esta metodología será también útil para analizar el posible efecto del cambio global en variables como la altura máxima diaria de la ola en una boya marina o la precipitación acumulada durante un mes en un observatorio.

Basándonos en estos resultados, se demuestra estadísticamente en este trabajo para el área mediterránea española el efecto del “stilling”, coincidimos con el suplemento especial del Boletín de la Sociedad Meteorológica Americana, “*State of the Climate in 2022*” [5], en el que se establece a nivel global que las velocidades máximas diarias de viento están disminuyendo.

Este trabajo tendrá continuidad en el marco del grupo de investigación de Métodos Estadísticos dedicado al análisis del cambio climático. Se utilizará la metodología para analizar las tendencias en observatorios de la España peninsular, con características climáticas muy diferentes en el interior o en la costa atlántica.

Por otra parte, el marco de estimación de modelos de regresión con distribución GEV permitirá abordar el estudio de atribuciones de dichas tendencias a la situación atmosférica, representada a partir de variables de la atmósfera libre, disponibles en el reanálisis ERA5 del European Centre for Medium-Range Weather Forecasts.



# Bibliografía

- [1] COLES, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- [2] BEDOYA-VALESTT, S., C. AZORIN-MOLINA, L. GIMENO, J. A. GUIJARRO, V. J. SANCHEZ-MORCILLO, E. AGUILAR, M. BRUNET, *Opposite trends of sea-breeze speeds and gusts in Eastern Spain, 1961–2019*, <https://link.springer.com/article/10.1007/s00382-022-06473-0>.
- [3] GILLELAND, E., R. W. KATZ, *extRemes 2.0: An Extreme Value Analysis Package in R*, <https://www.jstatsoft.org/article/view/v072i08>.
- [4] SOUKISSIAN, T., H., C. TSALIS, *The effect of the generalized extreme value distribution parameter estimation methods in extreme wind speed prediction*, <https://link.springer.com/article/10.1007/s11069-015-1800-0>.
- [5] BLUNDEN, J., T. BOYER, E. BARTOW-GILLIES, *Eds., 2023: “State of the Climate in 2022”*. *Bull. Amer. Meteor. Soc.*, 104 (9), Si–S501, <https://journals.ametsoc.org/view/journals/bams/104/9/2023BAMSSStateoftheClimate.1.xml>.
- [6] R CORE TEAM (2021), *R: A language and environment for statistical computing*. *R Foundation for Statistical Computing, Vienna, Austria.*, <https://www.R-project.org/>.
- [7] RSTUDIO TEAM (2020), *RStudio: Integrated Development for R*. *RStudio, PBC, Boston, MA*, <http://www.rstudio.com/>.
- [8] GILLELAND, E., R. W. KATZ, *extRemes 2.0: An Extreme Value Analysis Package in R*, <https://staff.ral.ucar.edu/ericg/extRemes/>.
- [9] ZHU, H., *kableExtra: Construct Complex Table with ‘kable’ and Pipe Syntax*, <http://haozhu233.github.io/kableExtra/>, <https://github.com/haozhu233/kableExtra>.





# ANEXO



## A.A. Conceptos teóricos

### A.A.1. Modelos asintóticos para mínimos

Algunas aplicaciones requieren modelos para observaciones extremadamente pequeñas, en lugar de extremadamente grandes. Considerando  $\tilde{M}_n = \min\{X_1, \dots, X_n\}$  y suponiendo que los  $X_i$  son independientes e idénticamente distribuidos, se aplican a  $\tilde{M}_n$  argumentos análogos a los aplicados a  $M_n$ , lo que conduce a una distribución límite de una variable convenientemente reescalada.

Los resultados son también inmediatos a los correspondientes de  $M_n$ . Dejando  $Y_i = -X_i$  para  $i = 1, \dots, n$ , el cambio de signo significa que valores pequeños de  $X_i$  corresponden a valores grandes de  $Y_i$ . Así, si  $\tilde{M}_n = \min\{X_1, \dots, X_n\}$  y  $M_n = \max\{Y_1, \dots, Y_n\}$ , entonces  $\tilde{M}_n = -M_n$ . Por lo tanto, para  $n$  grande,

$$\Pr\{\tilde{M}_n \leq z\} \approx 1 - \exp\left\{-\left[1 - \xi\left(\frac{z - \tilde{\mu}}{\sigma}\right)^{-1/\xi}\right]\right\},$$

en  $\{z : 1 - \xi(z - \tilde{\mu})/\sigma > 0\}$ , donde  $\tilde{\mu} = -\mu$ .

Esta distribución es la **distribución GEV para mínimos**. Podemos enunciar el resultado formalmente como un teorema análogo al **Teorema 3.1.1** para máximos.

**Teorema 3.3** Si existen secuencias de constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tales que

$$\Pr\left\{\frac{\tilde{M}_n - b_n}{a_n} \leq z\right\} \longrightarrow \tilde{G}(z) \quad \text{cuando } n \longrightarrow \infty$$

para una función de distribución no degenerada  $\tilde{G}$ , entonces  $\tilde{G}$  es un miembro de la familia GEV

$$\tilde{G}(z) = 1 - \exp\left\{-\left[1 - \xi\left(\frac{z - \tilde{\mu}}{\sigma}\right)^{-1/\xi}\right]\right\},$$

definida en  $\{z : 1 - \xi(z - \tilde{\mu})/\sigma > 0\}$ , donde  $-\infty < \mu < \infty$ ,  $\tilde{\sigma} > 0$  y  $-\infty < \xi < \infty$ .  $\square$

En las situaciones en las que conviene modelizar los mínimos en bloque, se puede aplicar directamente la distribución GEV para los mínimos. Una alternativa es explotar la dualidad entre las distribuciones para máximos y mínimos.



## A.B. Tendencias opuestas de las velocidades y ráfagas de las brisas marinas en el este de España, 1961-2019

La mayoría de los estudios sobre la variabilidad del viento han profundizado en los fenómenos de calma versus inversión a escala global y regional, mientras que los cambios a largo plazo en los vientos a escala local, como la brisa marina (SB en adelante), representan una brecha de conocimiento en la investigación climática.

En este apartado vamos a ver un estudio de variabilidad del viento, más concretamente sobre las SB en las estaciones costeras. Se ha desarrollado un método automatizado robusto para la identificación de días SB, y mediante las observaciones de viento homogeneizadas de 16 estaciones en el este de España, se analizará la variabilidad multidecenal y las tendencias en las velocidades, ráfagas y ocurrencia de SB para 1961 – 2019. El hallazgo principal serán las tendencias opuestas y variabilidad desacoplada de las velocidades y ráfagas de SB.

### A.B.1. Introducción

Desde una perspectiva global, los vientos costeros y de mar abierto soplan con más fuerza en el hemisferio sur y en los sistemas de surgencias, mientras que el aquietamiento terrestre (disminución de los vientos superficiales) continúa en latitudes medias a altas del hemisferio norte. Entre las posibles causas discutidas para los fenómenos de calma e inversión atmosférica, los cambios en el uso de la tierra y la rugosidad de la superficie y las oscilaciones internas decenales océano-atmósfera son los impulsores más probables, cambiando además de manera diferente en signo y magnitud a escala estacional, latitudinal, con diferencias entre estaciones costeras, de interior y de montaña.

En el área mediterránea, donde las condiciones climáticas favorecen el desarrollo de SB, las tendencias varían localmente, evidenciando que los cambios de SB se basan en la ubicación, diferente orientación de la costa y al terreno complejo, cambios en el uso de la tierra, circulación atmosférica a gran escala, aumento de la temperatura de la superficie u oscilaciones océano-atmósfera entre otros. La investigación multidecenal sobre SB es necesaria, ya que estos vientos tienen un gran impacto en amplias esferas socioeconómicas y ambientales, por tanto es de gran interés científico profundizar en el conocimiento de las tendencias de largo plazo de la SB en el marco del fenómeno de quietud versus inversión.

El objetivo de este estudio es analizar por primera vez las tendencias a largo plazo, la variabilidad multidecenal y los posibles impulsores de los cambios de velocidad del viento cerca de la superficie (media y ráfagas) en días dominados por circulaciones de SB en el Este de España durante 1961 – 2019. Más concretamente, los objetivos son:

- Desarrollar un método de selección robusto y automático para crear la primera base de datos de SB a largo plazo a partir de series de viento fiables y homogeneizadas.
- Avanzar en el estudio de los cambios estacionales en los fenómenos de quietud-inversión desde una perspectiva del tipo de viento (es decir, circulaciones SB) y tanto para velocidades medias como de ráfagas.
- Mejorar la comprensión de la influencia de la circulación atmosférica a gran escala sobre la variabilidad y las tendencias de la SB en un clima más cálido.

### A.B.2. Datos y métodos

#### Región de estudio

Definimos el este de España como el área que comprende la costa mediterránea de España, caracterizada por una topografía compleja al tener llanuras costeras, deltas, desembocaduras

de ríos, y valles amenizados por cadenas montañosas del interior como los sistemas Bético, Prebético e Ibérico. El clima es típicamente mediterráneo con inviernos suaves y veranos secos y calurosos, modulados por el efecto refrescante de las SB, que representan la circulación de viento más dominante en verano, pudiendo desarrollarse incluso durante el semestre frío al ser las condiciones atmosféricas estables frecuentes durante todo el año.

Dividimos el este de España en cinco regiones según la orientación de la costa, que influye fuertemente en la dirección dominante de SB: la Costa Sur, la Costa Sureste, la Costa Este, la Costa Nororiental y las Islas Baleares, visibles en la figura 3.

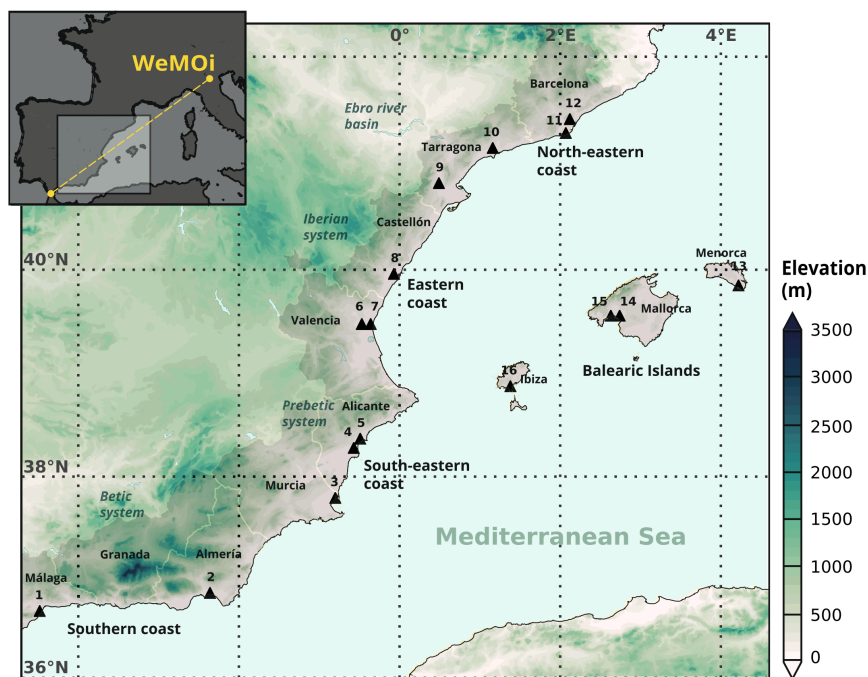


Figura 3: Mapa del terreno del Este de España con la ubicación de las 16 estaciones meteorológicas. La imagen superior izquierda insertada muestra el transecto del WeMOi utilizado para detectar eventos SB.

### Datos de velocidad del viento observados

La Agencia Estatal de Meteorológica Española (AEMET) proporcionó los datos diarios de velocidad media del viento promediados a partir de mediciones estándar de altura de 10 m y las ráfagas de viento máximas diarias (en 3 s) para las 16 estaciones del estudio.

Dichas estaciones no exceden los 18 km de distancia a la línea costera, ya que la penetración tierra adentro de la SB podría verse reducida en invierno o inhibirse por flujos sinópticos marinos. Además, cinco de las estaciones están ubicadas en entornos urbanos, mientras que el resto están ubicadas en sitios bien expuestos en aeropuertos, lo que reduce la frecuencia de posibles faltas de homogeneidad debido a cambios en la urbanización circundante.

### Homogeneización de los datos de velocidad del viento

Las series de velocidad del viento se crearon aplicando un robusto protocolo de control de calidad y homogeneización. Las reconstrucciones de las series de velocidad del viento se realizaron verificando discontinuidades, debido a la instalación de estaciones meteorológicas automáticas,

y todas las estaciones fueron homogeneizadas para evitar perturbaciones en las tendencias climáticas debido a puntos de ruptura no climáticos en las series de vientos. Además, también se utilizaron estaciones vecinas con alta correlación para detectar faltas de homogeneidad. El procedimiento no detectó ningún valor claramente atípico, pero corrigió 68 y 76 puntos de ruptura para velocidades de SB y ráfagas respectivamente.

### Identificación y validación de días de brisa marina

Los algoritmos de selección de SB consisten en detectar eventos pasados a partir de criterios bien definidos, como ciertos umbrales de viento o contraste térmico entre aire tierra-mar. Sin embargo, no existe un método universal para detectar con precisión los días SB. A su vez, la identificación de SB depende principalmente de los criterios utilizados en cada sitio de estudio, la complejidad física del área y la disponibilidad de datos.

En este estudio, se propone un método automatizado robusto para identificar posibles días de SB utilizando variables alternativas que pueden usarse en todo el mundo en cualquier región costera con pocos ajustes. El algoritmo de selección automatizada se basa en seis filtros alternativos, clasificados en condiciones sinópticas y de escala local. El primer grupo principal (filtros 1 a 3) emplea patrones de teleconexión regionales, tipos de clima sinóptico y vientos geostróficos como criterios alternativos para rechazar días con perturbaciones sinópticas a gran escala para que se desarrolle SB. El segundo grupo (filtros 4 a 6) utiliza criterios objetivos en el clima local para confirmar el paso de SB (por vientos terrestres y ráfagas de viento débiles) y para asegurar eventos de SB puros (días sin precipitación).

Distinguimos entre días de SB a nivel de estación (número exacto de eventos identificados para cada estación meteorológica), nivel costero (abarca todos los episodios de SB presentados en una misma región costera) y nivel regional (todos los días de SB identificados para todas las estaciones). Los filtros son los siguientes:

- **Filtro 1:** Condiciones débiles de presión superficial sobre la cuenca del Mediterráneo occidental.
- **Filtro 2:** Ocho tipos de clima (A, E, NE, AE, ANE, C, N y SE) filtrados como condiciones sinópticas favorables para SB.
- **Filtro 3:** Los vientos geostróficos débiles ( $\leq 12 \text{ m s}^{-1}$ ) excluyen los días dominados por fuertes flujos sinópticos, que perturban el desarrollo de la SB.
- **Filtro 4:** Los vientos del mar fueron seleccionados en base a la orientación de la costa y la dirección del viento a las 12 : 00 UTC. Además, para las Islas Baleares, permitimos todas las direcciones del viento debido a la complejidad de definir la orientación de la costa, y éstas pueden verse afectadas por los flujos de SB desde todas las direcciones. Se calcula la dirección del viento a partir de los componentes del viento zonal (u) y meridional (v) a 10 m de altura a las 12 : 00 UTC.
- **Filtro 5:** Se requirieron ráfagas de viento diarias ( $\leq 13,9 \text{ m s}^{-1}$ ) para excluir los flujos terrestres moderados-fuertes.
- **Filtro 6:** Se necesitó precipitación total diaria ( $\leq 0,1 \text{ mm d}^{-1}$ ) para rechazar los días con inestabilidad atmosférica.

### Análisis de tendencias y estadísticas

Se estimaron las tendencias a largo plazo de las velocidades, ráfagas y ocurrencia de SB para el periodo 1961–2019. Para ello se crearon series temporales diarias de velocidad y ráfagas de viento para los días SB en cada estación meteorológica, para obtener series de recurrencia mensuales

promedio y mensuales. Seguidamente se realizaron los promedios ponderados agrupando todos los días SB encontrados en una misma región para crear series regionales y costeras, para calcular las anomalías con respecto a 1981 – 2010 y así asegurar que ninguna estación meteorológica dominará la serie promediada. Se estimaron las tendencias con una regresión lineal ponderada en metros por segundo por década, donde la significancia estadística fue evaluada desde una perspectiva de “proceso e importancia” utilizando tres umbrales de nivel  $p$  definidos como:

- Altamente significativos ( $p < 0,05$ ).
- Significativo ( $p < 0,10$ ).
- No significativo ( $p > 0,10$ ).

También se evaluó la presencia de la tendencia realizando tendencias corridas que detectan subperiodos de tendencias más debilitadas o más fuertes. Además, se ilustró la variedad multi-decanal de SB y el coeficiente del valor de determinación ( $R^2$ ) para estimar la varianza entre velocidades de SB y ráfagas. Todos los cálculos se realizaron en tres escalas de tiempo diferentes: anual, mensual y estaciones boreales. Por último, para explorar la influencia de la circulación atmosférica a gran escala en las tendencias y variabilidad de la SB, correlacionamos las velocidades, ráfagas y días ponderados de la SB con, por un lado los índices de circulación atmosférica a escala anual y estacional, y por otro con la frecuencia de los 26 tipos de clima JC a escala mensual, agrupados de la siguiente forma:

- Tipos de flujo direccional: N, NE, E, SE, S, SW, W y NW.
- Tipos anticiclónicos: A, y ciclónicos: C.
- Tipos híbridos: AN, ANE, AE, ASE, AS, ASW, AW, ANW, CN, CNE, CE, CSE, CS, CSW, CW y CNW.

### Índices de circulación atmosférica

Se seleccionaron cuatro índices de circulación atmosférica para evaluar su relación a través del coeficiente de correlación de Pearson ( $r$ ) con la variabilidad de SB observada durante el estudio de 58 años:

- El diario WeMOI, basado en la diferencia regional de presión a nivel del mar entre el dipolo barométrico San Fernando (Cádiz) y Padua (Italia) para detectar posibles pasajes de SB para el este de España.
- El MOI.
- El NAOI.
- El SNAOI, calculado promediando los valores del NAOI de Julio a Agosto.

Estos índices impulsan la mayor parte de la variabilidad climática en el este de España.

### A.B.3. Resultados

#### Climatología SB

En general, se observa una marcada estacionalidad de la SB, con mayor ocurrencia en verano, y la baja ocurrencia de SB en invierno parece aumentar ligeramente en los últimos años.

La climatología anual está dominada por la temporada primavera-verano. Las regiones nor-oriental y meridional exhibieron los valores más grandes en todas las escalas, mientras que los más pequeños se produjeron en la región oriental. A escala intraanual, el verano exhibió las



mayores velocidades de SB y valores de ráfagas, mientras que en invierno se encontraron fuertes diferencias entre las regiones, siendo estos valores mayores en las estaciones de aeropuerto (zonas rurales), que en las ciudades, debido al efecto de la urbanización en todos los parámetros estadísticos.

### **Tendencias en las velocidades, ráfagas y ocurrencia de la SB**

Se encontraron tendencias opuestas generalizadas entre las velocidades de SB y las ráfagas. Para 1961 – 2019, las velocidades regionales de SB mostraron disminuciones significativas en todas las escalas de tiempo, pero las ráfagas de SB presentaron tendencias anuales crecientes no estacionales. Las tendencias son positivas en invierno (no significativas) y otoño (significativas), y aparecen pendientes negativas opuestas en primavera y verano. Es notable la reciente desaceleración significativa de las velocidades de SB en otoño-invierno después de un largo periodo de estabilidad no significativa o aumentos débiles.

En los últimos años se ha observado un debilitamiento reciente de las ráfagas SB invernales, mientras que por el contrario, las de primavera-verano comenzaron a reforzarse. Se encontró una marcada desaceleración en las velocidades de SB durante todo el año en las regiones costeras, sin embargo, las regiones del este y nordeste mostraron aumentos significativos en otoño-invierno, aunque comenzaron a debilitarse significativamente en los últimos años. Las tendencias de las ráfagas de SB fueron en general heterogéneas y no significativas, pero pequeños aumentos dominaron las regiones oriental y nororiental en invierno y otoño, mientras que, por el contrario, la disminución fue evidente en primavera para gran parte de las estaciones y persistente en Baleares durante casi todas las estaciones.

En particular, algunas ciudades evidenciaron tendencias opuestas entre aeropuertos y estaciones urbanas en diferentes escalas temporales para ambos parámetros. La ocurrencia regional de SB mostró tendencias positivas no significativas anualmente, destacando la importante tendencia al alza en invierno y la tendencia a la baja en verano. El aumento de la ocurrencia de SB es aún más evidente a nivel costero para la región sur y este con magnitudes de cambio anual, con una tendencia positiva notable y significativa en invierno para todos. Por último, cabe destacar la tendencia al alza en las estaciones de aeropuerto frente al descenso en las estaciones urbanas dentro de una misma ciudad, característica clara del efecto de la rugosidad superficial local.

### **Variabilidad multidecadal de las velocidades, ráfagas y días de la SB**

En general, el filtro gaussiano de paso bajo de 15 años reveló una variabilidad desacoplada entre velocidades y ráfagas de SB a escala anual, así como en verano. Para estas escalas de tiempo, las anomalías para las velocidades de SB disminuyeron y las ráfagas de SB aumentaron con cambios simultáneos para ambos parámetros alrededor de la década de 1990. La variabilidad intraanual más fuerte de anomalías ocurrió en invierno-otoño, especialmente para las ráfagas SB, y la más débil en verano. El invierno también presenta una fuerte correlación positiva entre las velocidades de SB y la variabilidad de las ráfagas, lo mismo para primavera y otoño. Para la ocurrencia de SB, es destacable la creciente variabilidad observada en invierno, y la tendencia decreciente en verano.

### **Influencia de la circulación atmosférica a gran escala**

Para los 26 tipos de clima JC, las correlaciones positivas más altas ocurrieron para los días SB, donde el régimen A mejoró la ocurrencia de eventos SB regionalmente desde otoño hasta primavera, seguido de los tipos E, NE y AE, los cuales ejercieron una fuerte influencia a lo largo del año. A nivel costero, la ocurrencia de SB en verano está claramente inhibida por los vientos del oeste, aunque una frecuente advección anómala del este puede potenciarla. Los resultados más notables para las velocidades de SB y las ráfagas fueron su significativa relación negativa

con el régimen A, lo que denota que una atmósfera estable puede debilitar las circulaciones de SB.

#### **A.B.4. Discusión**

Este estudio investigó por primera vez los cambios a largo plazo en las velocidades, ráfagas y ocurrencia de SB en el este de España durante el periodo 1961 – 2019, proponiendo un método automatizado para identificar eventos de SB pasados mediante la aplicación de seis criterios de prueba alternativos, con el objetivo de crear la base de datos disponible más larga (58 años) de posibles episodios de SB en toda la costa mediterránea española y proporcionar nuevos conocimientos sobre el clima multidecenal y la variabilidad de este viento local.

Los principales hallazgos revelan una nueva evidencia sobre los cambios en la SB en el Este de España, una región en la que la SB representa la circulación de viento local más frecuente a lo largo del año. Además, se muestra una variabilidad desacoplada y tendencias entre velocidades de SB y ráfagas, y se llega a la conclusión de que comprender los cambios de SB es un desafío porque los conductores pueden diferir entre velocidades medias y de ráfagas.

#### **Aumento de ocurrencia de SB por el efecto anticiclónico actividad y otras dinámicas relacionadas a gran escala**

La tendencia de la SB en el Este de España sugiere que se está convirtiendo en un fenómeno menos común en verano, pero en discordancia con el hipotético aumento inducido por el aumento de la temperatura en el aire. Por el contrario, el aumento de los días de SB en invierno concuerda con el hallazgo reciente de una expansión de las Altas Azores en el siglo XX, en la que las circulaciones son significativamente más frecuentes por el Mediterráneo occidental y la Península Ibérica en invierno, pudiendo estar detrás del aumento de días SB durante los meses fríos.

Este patrón también se proyecta para el futuro y es consistente con el calentamiento impulsado antropogénicamente, por lo que podríamos esperar un aumento de la aparición de SB en invierno durante el siglo XXI. El aumento de la ocurrencia de SB también está fuertemente relacionado con el E, mientras que otros regímenes como los tipos AE y NE lo favorecen en parte. El patrón más evidente encontrado estacionalmente es el efecto negativo de los vientos del oeste y del noroeste al inhibir la ocurrencia del verano.

El siguiente paso lógico de esta investigación es evaluar el impacto de los cambios de los flujos sinópticos en las características de la SB. Las circulaciones anticiclónicas invernales sobre la región (y por tanto la actividad SB) pueden verse favorecidas por el NAOI+. La fuerte correlación encontrada entre los días NAOI y SB proporcionan información plausible sobre la relación dominante entre NAOI+, régimen A y el aumento de la actividad SB en la región en invierno. Dado que NAO y MO han demostrado estar fuertemente vinculados en invierno, es posible que los modos acoplados de circulación atmosférica o su interacción con mecanismos locales que tienen lugar en los meses cálidos podría tener un mayor peso a la hora de explicar la aparición del SB de verano.

#### **Incertidumbres en los mecanismos que impulsan las velocidades y ráfagas de SB desacopladas**

El calentamiento global está mejorando la estabilidad en latitudes medias debido a hundimientos que pueden causar circulaciones de SB más frecuentes pero más débiles, y esto es consistente con los hallazgos del estudio de una mayor actividad de SB pero velocidades de SB debilitadas a lo largo del año. Los resultados sugieren que las circulaciones anticiclónicas más frecuentes están en parte detrás de la disminución de las velocidades y ráfagas de SB durante el invierno. Además, trabajos recientes sugieren que el debilitamiento de los vientos locales estivales

en Europa occidental se debe al efecto del aumento de la temperatura del aire y el calentamiento de los océanos al reducir el gradiente térmico tierra-mar.

Otros autores sugirieron el efecto de la expansión urbana y los patrones de riego en la disminución de las velocidades de SB, siendo mecanismos que pueden explicar aún más las tendencias locales (especialmente en otoño-invierno), así como las opuestas entre lo rural y lo urbano. Finalmente, la disminución en las velocidades de la SB es contraria a la hipótesis generalizada de que una SB reforzada provoca vientos costeros intensificados en verano, con la tendencia creciente de la velocidad del viento cerca de la superficie en verano reportada para España.

Para SB, la influencia de los vientos sinópticos a gran escala en ubicaciones específicas podría estar detrás de la tendencia desacoplada entre las velocidades de SB y las ráfagas, pero probablemente esté acoplada a factores locales, y mecanismos tales como la degradación de la tierra y topografía, el calentamiento de la temperatura del aire, el agotamiento del suelo, los flujos sinópticos terrestres y su interacción con la orografía (influyen fuertemente en los vientos extremos, por ejemplo, ráfagas y frentes de SB). El cambio climático está potenciado por los eventos extremos, y un análisis exhaustivo de los efectos locales y el calentamiento global está más allá del alcance de esta investigación, por lo que futuras investigaciones podrían ayudar a revelar las causas detrás de las tendencias opuestas en las velocidades y ráfagas de SB.

### Descripción general, posibles implicaciones y comentarios finales

Nunca se han informado en la literatura tendencias opuestas en las velocidades y ráfagas regionales de SB para el Este de España. Cabe mencionar que el fortalecimiento de las rachas de SB se corresponde con el aumento de las rachas de viento cercanas a la superficie en la Península Ibérica durante los meses cálidos para estaciones específicas. Además, la naturaleza local del SB hace que la comparación de tendencias entre regiones sea aún más compleja. Entre los mecanismos a gran escala investigados aquí, no se han encontrado los procesos detrás de las crecientes ráfagas de SB y su variabilidad de desacoplamiento con las velocidades de SB.

#### A.B.5. Conclusiones

Los principales hallazgos de las tendencias a largo plazo y la variabilidad de la aparición, velocidades y ráfagas de brisa marina en el Este de España para el periodo 1961 – 2019 son:

- La ocurrencia de SB aumentó anualmente y en invierno. Por el contrario, los días de SB en verano disminuyeron.
- Se observaron tendencias opuestas entre las velocidades de ráfagas con marcada estacionalidad, es decir, las velocidades SB disminuyeron significativamente en primavera y verano, mientras que las ráfagas de SB aumentaron en otoño, pero no significativamente en invierno.
- Se cree que las tendencias en la aparición de SB en invierno están determinadas por condiciones anticiclónicas anómalas frecuentes y el NAOI+ y MOI+, al tiempo que inhiben las velocidades de SB. La disminución de las velocidades de SB durante el invierno se debe principalmente a la influencia del régimen A, mientras que el refuerzo de las ráfagas de SB sigue siendo incomprendido.

Los futuros estudios de sensibilidad que utilicen modelos climáticos podrían ayudar a comprender mejor los factores que controlan su variabilidad en un clima cada vez más cálido.



## A.C. Indicadores AEMET

Información acerca de los indicadores de AEMET para el lugar:

- ◇ X6155A  $\implies$  Aeropuerto de Málaga
- ◇ X6325O  $\implies$  Aeropuerto de Almería
- ◇ X7031  $\implies$  Aeropuerto de San Javier
- ◇ X8019  $\implies$  Aeropuerto de Alicante
- ◇ X8025  $\implies$  Alicante Ciudad Jardín
- ◇ X8414A  $\implies$  Aeropuerto de Valencia
- ◇ X8416  $\implies$  Valencia Viveros
- ◇ X8500A  $\implies$  Castellón - Almassora
- ◇ X9981A  $\implies$  Tortosa
- ◇ X0016A  $\implies$  Aeropuerto de Reus
- ◇ X76  $\implies$  Aeropuerto de Barcelona
- ◇ X0200E  $\implies$  Barcelona Fabra
- ◇ B893  $\implies$  Aeropuerto de Menorca
- ◇ B278  $\implies$  Aeropuerto de Palma de Mallorca
- ◇ B228  $\implies$  Puerto de Palma de Mallorca
- ◇ B954  $\implies$  Aeropuerto de Ibiza



## A.D. Análisis exploratorio

A continuación aparecen tablas y figuras adicionales del análisis exploratorio:

### A.D.1. Aeropuerto de Reus

Tabla descriptiva de la velocidad en el Aeropuerto de Reus

Mes	Media	Desv. típica	Mediana	Coef. de variación	Coef. de asimetría
Enero	44.51996	25.25102	38	0.5671843	0.7746169
Febrero	45.07743	24.38915	39	0.5410501	0.7475575
Marzo	44.78513	21.89045	39	0.4887884	0.7928291
Abril	43.53616	19.81980	38	0.4552491	0.8379741
Mayo	38.78458	15.77760	34	0.4068009	0.9097544
Junio	35.78531	12.08755	34	0.3377796	0.4430951
Julio	36.28978	12.73889	34	0.3510324	0.5392408
Agosto	36.17387	13.54215	33	0.3743629	0.7031081
Septiembre	35.92881	14.38062	32	0.4002531	0.8196060
Octubre	36.85511	19.73116	31	0.5353711	0.8902333
Noviembre	40.72260	21.88746	35	0.5374770	0.7843668
Diciembre	43.14872	25.01526	36	0.5797453	0.8573224

Tabla 2: Descriptiva de la velocidad en Reus distinguiendo por meses compuesta por la media, desviación típica, mediana, y coeficientes de variación y asimetría.

### Media y mediana de la velocidad máxima por mes en el aeropuerto de Reus

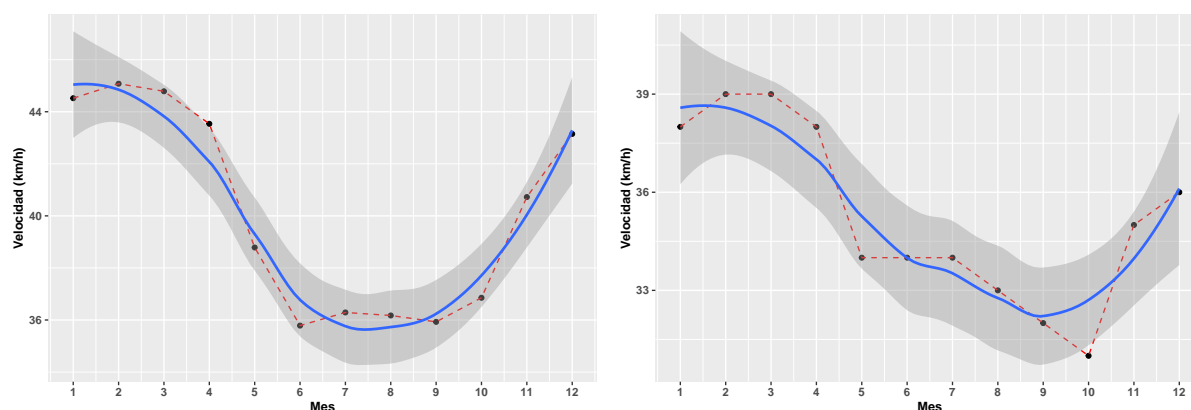


Figura 4: Media (izda.) y mediana (dcha.) de la velocidad máxima expresada en km/h para cada mes en Reus. En ambos casos las curvas se han ajustado con `geom_smooth` usando el método `loess` con bandas de confianza del 95 %.

La velocidad media ronda entre los 35 y 45 km/h, siendo mayor en los meses de invierno, y reduciéndose en los meses más calidos. Esto nos muestra que estamos ante unos datos con una marcada estacionalidad. Podemos ver que el mes con mayor velocidad media es Febrero, con un valor de aproximadamente 45 km/h, mientras que el valor más bajo se obtiene en Junio con un valor cercano a 35 km/h.

### Diferencia entre los valores de los cuantiles de ambos decenios

En la figura 5 vemos que la diferencia entre ambos periodos es que los valores de la mediana estan cercanos al 0 en todo el año salvo en los últimos meses del año. Los cuantiles encima de la mediana ya son negativos durante todo el año, siendo bastante negativos en el cuantil 90.

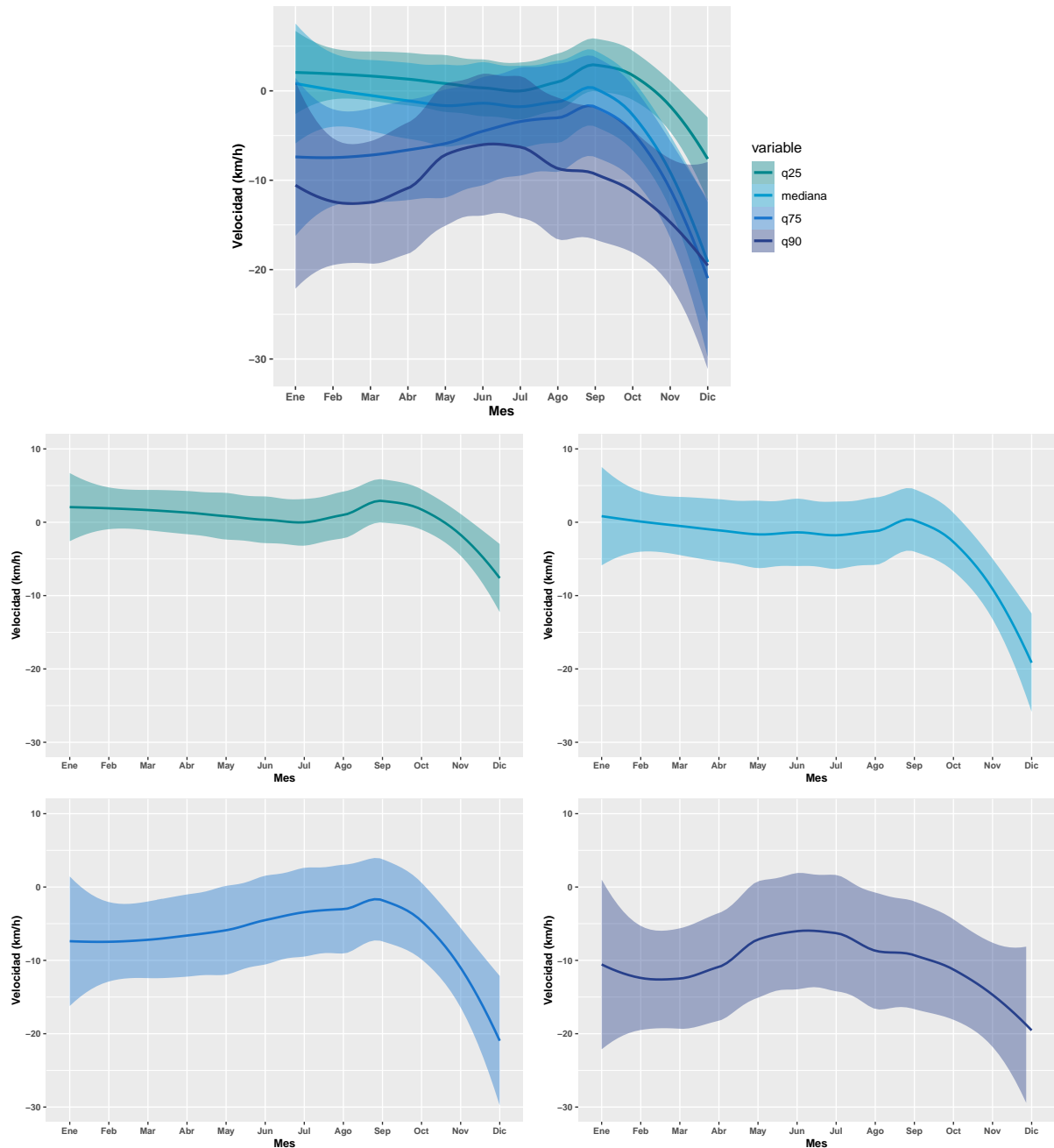


Figura 5: Valores resultantes de la diferencia de los cuantiles 25, 50, 75 y 90 de la velocidad máxima para cada mes en Reus entre los periodos 2010 – 2019 y 1961 – 1970. Las curvas se han ajustado con `geom_smooth` usando el método `loess` con bandas de confianza del 95 %.

### Coefficiente de variación y asimetría

En la figura 6 vemos el coeficiente de variación y de asimetría de la velocidad máxima para cada mes en Reus.



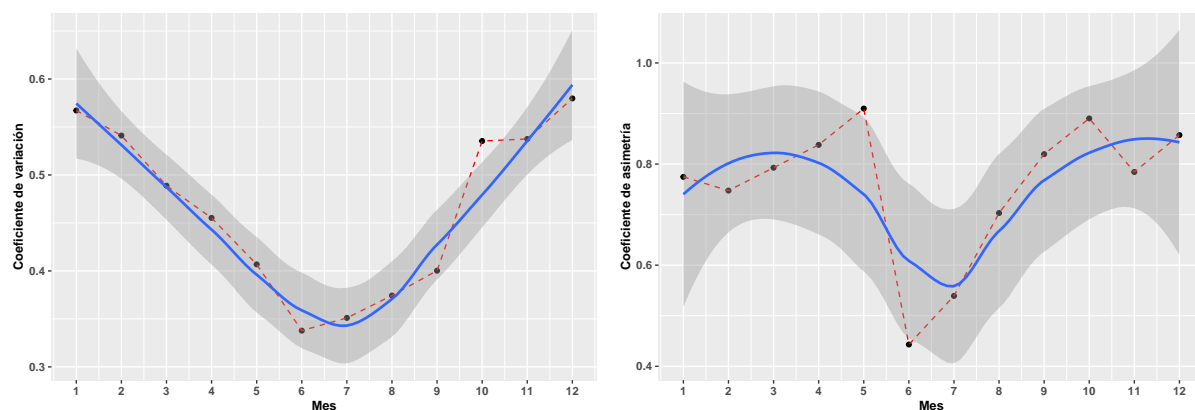


Figura 6: Coeficiente de variación (izda.) y coeficiente de asimetría (dcha.) de la velocidad máxima para cada mes en Reus. En ambos casos las curvas se han ajustado con `geom_smooth` usando el método `loess` con bandas de confianza del 95 %.

### Histograma y diagrama de cajas

La figura 7 representa el histograma de la velocidad en los meses de Enero y Julio, y el diagrama de cajas la velocidad en todos los meses del año. Para el histograma, a la vista de la gráfica podemos observar que la mayoría de los valores están entre los 25 y 50 km/h, y que en invierno hay valores elevados de velocidad que no se dan en verano.

Para el diagrama de cajas, observamos que los datos para los meses de invierno tienen un sesgo superior que para los meses de verano. Además, volvemos a corroborar que en invierno los valores de velocidad son superiores con respecto a los meses de verano, donde en éstos los valores que están entre los percentiles 25 y 75 abarcan menos cantidad no atípica, siendo en invierno esta cantidad mucho mayor a la vista de la gráfica. Por otro lado observamos valores típicos en prácticamente todos los meses del año, siendo destacable los valores inferiores en los meses de verano.

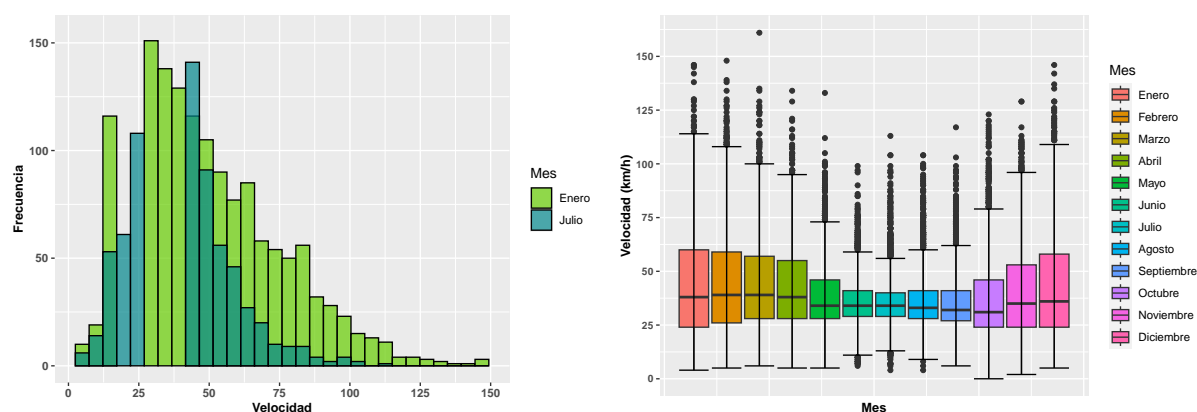


Figura 7: Izda., histograma de la velocidad en Reus durante los meses de Enero y Julio con los datos crudos, y la orden `geom_histogram`, filtrando los datos según sea Enero o Julio. Dcha., diagrama de cajas de la velocidad en Reus durante todos los meses del año con la orden `geom_boxplot`.

### Gráficas de la velocidad máxima en el Aeropuerto de Reus cada día del año

La figura 8 representa la descriptiva de la velocidad máxima en el Aeropuerto de Reus para cada día del año. Para facilitar la comprensión de éstas pondremos el comienzo de cada mes.

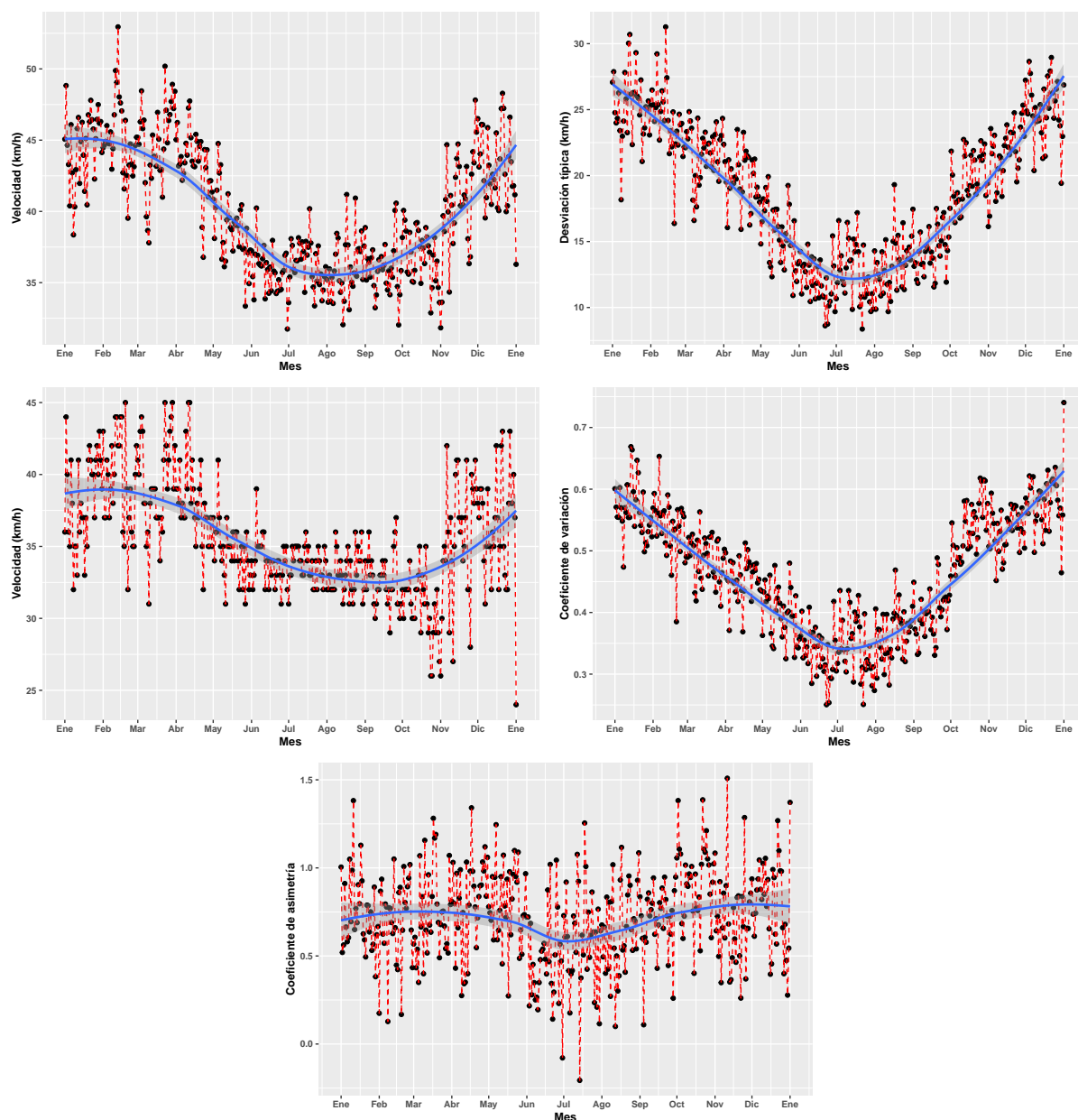


Figura 8: Media, desviación típica, mediana, y coeficientes de variación y asimetría de la velocidad máxima a lo largo de los días del año en el Aeropuerto de Reus diferenciando en meses.

## A.D.2. Gráficas de la velocidad máxima y desviación típica por zonas

⊗ **COSTA SUDORIENTAL**  $\Rightarrow$  Sistema Prebético y relieve costero irregular, formada por el Aeropuerto de San Javier (X7031), Aeropuerto de Alicante (X8019) y Alicante Ciudad Jardín (X8025). Figura 9.

⊗ **COSTA SUR**  $\Rightarrow$  Sistema Bético con el pico más alto de la sierra (Mulhacen), formada por el Aeropuerto de Málaga (X6155A) y Aeropuerto de Almería (X6325O). Figura 10.

⊗ **ISLAS BALEARES**  $\Rightarrow$  Conjunto de islas con una orografía bastante compleja, formada por el Aeropuerto de Menorca (B893), Aeropuerto de Palma de Mallorca (B278), Puerto de Palma de Mallorca (B228) y Aeropuerto de Ibiza (B954). Figura 11.

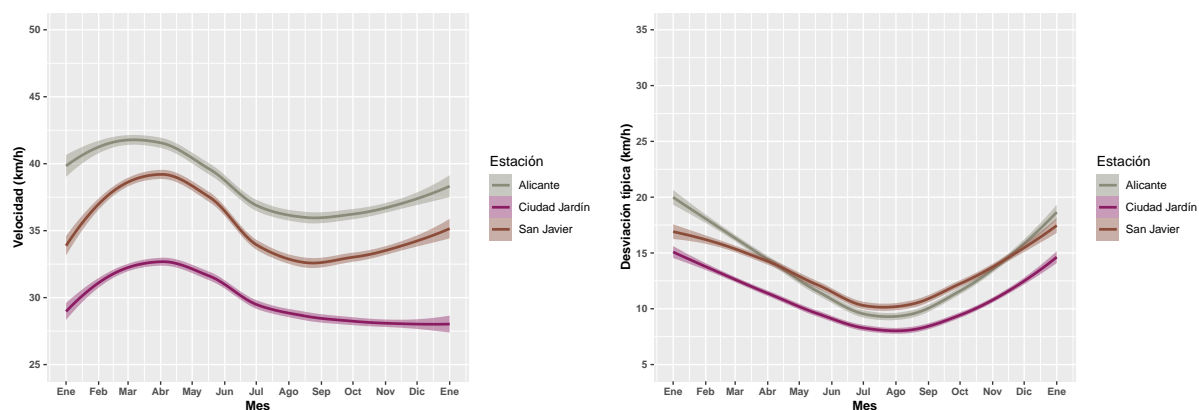


Figura 9: Media (izda.) y desviación típica (dcha.) de la velocidad máxima expresada en km/h para cada mes en las estaciones de la Costa Sudoriental. En ambos casos las curvas se han ajustado con `geom_smooth` usando el método `loess` con bandas de confianza del 95 %.

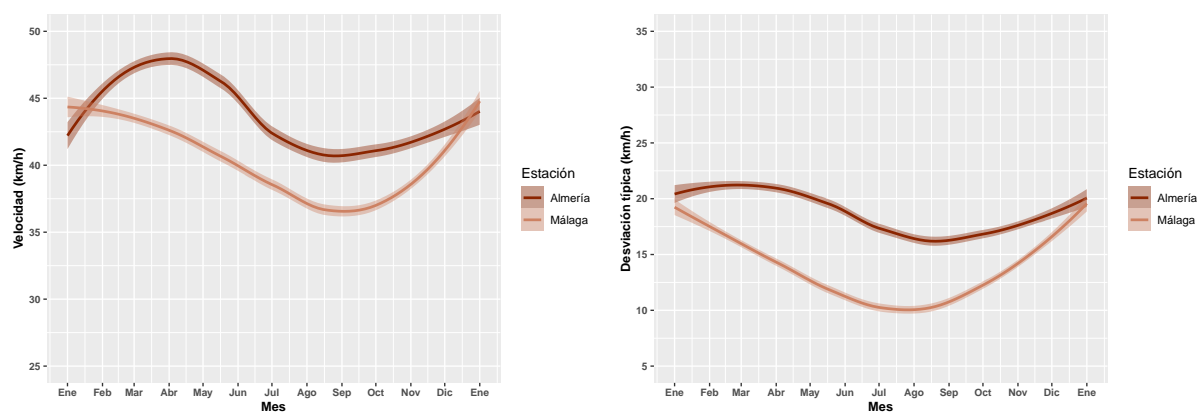


Figura 10: Media (izda.) y desviación típica (dcha.) de la velocidad máxima expresada en km/h para cada mes en las estaciones de la Costa Sur. En ambos casos las curvas se han ajustado con `geom_smooth` usando el método `loess` con bandas de confianza del 95 %.

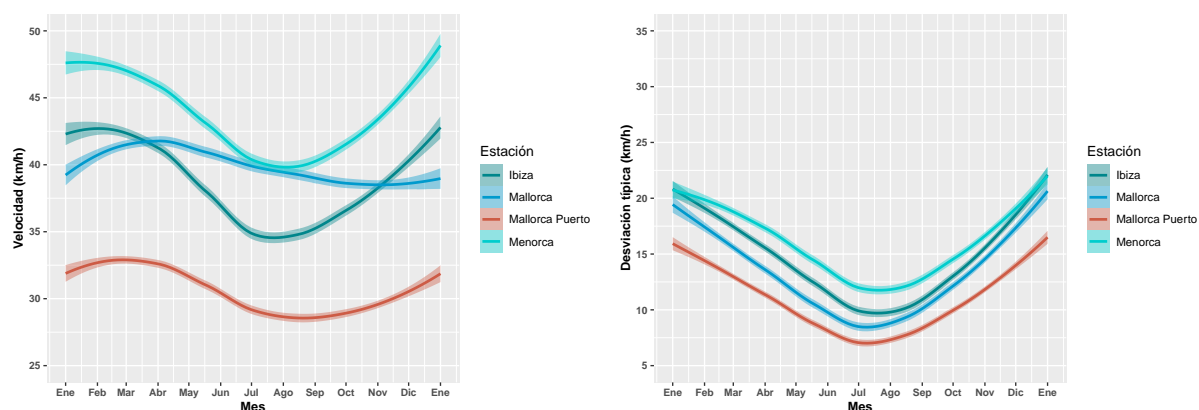


Figura 11: Media (izda.) y desviación típica (dcha.) de la velocidad máxima expresada en km/h para cada mes en las estaciones de las Islas Baleares. En ambos casos las curvas se han ajustado con `geom_smooth` usando el método `loess` con bandas de confianza del 95 %.



## A.E. Modelización inicial

### A.E.1. Aeropuerto de Reus

Tabla comparativa de la distribución Gumbel y tendencia en Reus por meses

Mes	AIC - S	AIC - G	$\mu$ - S	$\mu$ - G	$\sigma$ - S	$\sigma$ - G	$\xi$ - S	IC $\xi <$	IC $\xi >$
Enero	16519.81	16581.14	31.35	33.17	16.85	18.52	0.19	0.14	0.24
Febrero	14917.70	14962.08	32.80	34.21	16.55	17.78	0.15	0.11	0.20
Marzo	15997.34	16028.02	34.15	35.07	15.30	16.02	0.11	0.07	0.15
Abril	15207.77	15218.73	34.10	34.64	14.49	14.88	0.07	0.03	0.11
Mayo	14814.54	14818.32	31.66	31.88	11.59	11.70	0.04	0.01	0.07
Junio	13669.70	13710.48	30.72	30.26	10.40	10.37	-0.08	-0.11	-0.06
Julio	14129.99	14137.40	30.93	30.74	10.15	10.13	-0.04	-0.06	-0.01
<b>Agosto</b>	14258.55	14256.56	30.32	30.33	10.22	10.23	<b>0.00</b>	<b>-0.02</b>	<b>0.03</b>
Septiembre	13946.36	13950.34	29.57	29.76	10.42	10.51	0.03	0.01	0.06
Octubre	15380.41	15457.49	27.46	28.45	12.68	13.48	0.14	0.11	0.18
Noviembre	15503.36	15535.79	29.89	30.93	15.26	16.12	0.12	0.08	0.17
Diciembre	16397.39	16494.27	29.85	32.01	15.88	17.95	0.24	0.19	0.29

Tabla 3: Valores AIC de ambos modelos, valores de los parámetros de localización  $\mu$ , escala  $\sigma$  y forma  $\xi$  según el modelo y los intervalos de confianza asociados al valor del parámetro de forma en el modelo sin distribución Gumbel (S). Los intervalos de confianza de  $\xi$  que contienen al 0 están marcados en negrita junto con el valor de  $\xi$  y el mes correspondiente.

Mes	$\mu_1$	p-valor	NDR-10		Cambio	NDR-50		Cambio
			1961	2019		1961	2019	
Enero	0.04	1.00	77.85	80.37	2.53	130.73	133.26	2.53
Febrero	0.03	0.00	75.68	77.58	1.90	120.28	122.18	1.90
Marzo	0.05	0.00	71.82	74.88	3.06	108.77	111.84	3.06
Abril	0.05	0.00	68.13	70.87	2.75	99.00	101.75	2.75
Mayo	0.05	0.00	57.30	60.43	3.13	79.80	82.93	3.13
Junio	0.05	0.00	50.71	53.70	2.99	64.76	67.74	2.99
Julio	0.09	0.00	50.01	55.44	5.43	65.93	71.36	5.43
Agosto	0.07	0.00	51.20	55.51	4.31	69.03	73.34	4.31
Septiembre	0.11	0.00	50.72	56.88	6.16	71.01	77.17	6.16
Octubre	0.12	0.00	57.74	64.45	6.71	92.89	99.59	6.71
Noviembre	0.06	0.00	68.02	71.45	3.44	106.93	110.37	3.44
Diciembre	-0.02	0.31	77.53	76.39	-1.15	131.25	130.11	-1.15

Tabla 4: Valor de  $\mu_1$  del modelo GEV con `location.fun = ~ year`, p-valor del test entre este modelo (con tendencia  $\mu_1$ ) y el modelo inicial GEV ( $\mu_1 = 0$ ), niveles de retorno (NDR) a 10 y 50 años en 1961 y en 2019, es decir, el primer y último año de los datos que disponemos respectivamente y cambio entre los niveles de retorno entre ambos años.

En la tabla 3 hemos obtenido valores AIC prácticamente iguales en ambos modelos, pero por otra parte la variabilidad de los datos en los coeficientes  $\mu$  y  $\sigma$  reafirma la estacionalidad obtenida en el análisis exploratorio anterior. Además,  $\xi$  - S muestra variaciones de datos también, siendo más altos en meses de invierno, y más bajos en los meses de verano, hecho que implica estacionalidad, y por tanto los tres coeficientes muestran estacionalidad. Para el intervalo de confianza de  $\xi$ , vemos que solamente contiene al 0 el mes de Agosto.

En la tabla 4 observamos que  $\mu_1 > 0$  en todos los meses salvo en Diciembre, y tenemos un **p-valor** cercano a 0 salvo en Enero y Diciembre. Además los niveles de retorno muestran que estos serán superados por término medio una vez cada 10 y 50 años con valores mayores en 2019, luego en prácticamente todo el año hay tendencia de que haya un aumento en la velocidad del viento, con cambios de entre 1 y 7 km/h, siendo este aumento más acusado en los meses de verano, donde hemos obtenido las pendientes más positivas.

## A.E.2. Aeropuerto de Almería

Realizamos lo mismo para la estación de Almería.

Mes	AIC - S	AIC - G	$\mu$ - S	$\mu$ - G	$\sigma$ - S	$\sigma$ - G	$\xi$ - S	IC $\xi <$	IC $\xi >$
Enero	15763.65	15783.13	33.62	34.27	14.62	15.08	0.08	0.04	0.12
Febrero	14698.11	14700.82	36.90	37.31	16.48	16.78	0.05	0.00	0.09
Marzo	16181.05	16181.74	38.33	38.02	17.45	17.24	-0.03	-0.07	0.01
Abril	15595.56	15615.30	40.12	39.15	17.79	17.17	-0.10	-0.14	-0.06
Mayo	15975.98	15978.23	37.45	37.03	16.58	16.29	-0.05	-0.09	0.00
Junio	15201.33	15201.40	34.51	34.80	14.70	14.93	0.04	-0.01	0.09
Julio	15187.87	15186.07	33.06	33.14	12.91	12.97	0.01	-0.04	0.06
Agosto	15033.78	15034.08	33.00	33.24	12.22	12.41	0.04	-0.01	0.08
Septiembre	14707.52	14706.28	33.62	33.76	12.90	13.01	0.02	-0.03	0.07
Octubre	15452.61	15463.22	33.27	33.77	13.52	13.89	0.07	0.03	0.11
Noviembre	15093.75	15091.87	34.43	34.38	14.66	14.63	-0.01	-0.04	0.03
Diciembre	15623.46	15627.95	33.86	34.20	14.37	14.59	0.04	0.01	0.08

Tabla 5: Valores AIC y parámetros de localización  $\mu$ , escala  $\sigma$  y forma  $\xi$  de ambos modelos e intervalos de confianza asociados a  $\xi$  en el modelo sin distribución Gumbel (S).

Mes	$\mu_1$	p-valor	NDR-10		Cambio 10	NDR-50		Cambio 50
			1961	2019		1961	2019	
Enero	0.08	1	67.44	72.12	4.67	99.89	104.56	4.67
Febrero	0.07	0	74.01	78.09	4.08	106.60	110.68	4.08
Marzo	-0.04	1	77.17	75.04	-2.14	102.89	100.75	-2.14
Abril	-0.01	0	76.16	75.51	-0.65	97.46	96.81	-0.65
Mayo	0.00	1	72.82	72.78	-0.04	96.56	96.52	-0.04
Junio	0.00	0	68.96	68.95	-0.01	96.03	96.02	-0.01
Julio	-0.02	0	63.04	61.80	-1.24	84.76	83.52	-1.24
Agosto	-0.03	0	62.60	60.78	-1.82	84.85	83.03	-1.82
Septiembre	0.00	0	63.41	63.25	-0.15	86.06	85.91	-0.15
Octubre	-0.01	0	66.33	65.85	-0.47	93.65	93.18	-0.47
Noviembre	0.06	0	65.60	68.97	3.37	90.31	93.68	3.37
Diciembre	0.11	0	64.91	71.01	6.09	94.43	100.53	6.09

Tabla 6: Valor de  $\mu_1$  del modelo GEV con `location.fun = ~ year`, p-valor del test entre este modelo (con tendencia  $\mu_1$ ) y el modelo inicial GEV ( $\mu_1 = 0$ ), niveles de retorno (NDR) a 10 y 50 años en 1961 y en 2019, y cambio entre los niveles de retorno entre ambos años.

En la tabla 5 hemos obtenido valores AIC prácticamente iguales en ambos modelos, y menor variabilidad de los datos en los coeficientes  $\mu$  y  $\sigma$  que reafirma existencia de estacionalidad, menos notoria que en Reus. Además,  $\xi$  - S muestra pequeñas variaciones de datos también, que no siguen

un determinado patrón, con valores muy cercanos al 0, luego los tres coeficientes muestran estacionalidad, bastante menos marcada que en Reus. Es preferible quedarse con modelo sin distribución Gumbel, ya que es un modelo menos rígido.

En la tabla 6 observamos valores de  $\mu_1$  con pendientes muy suaves e incluso sin pendiente en meses, destacando valores positivos más altos en invierno y valores negativos en los meses de primavera y verano, y tenemos un **p-valor** prácticamente todo el año salvo en Enero, Marzo y Mayo. Además los niveles de retorno muestran una tendencia de que haya un aumento, disminución e incluso sin cambios en la velocidad del viento, con cambios de entre 0 y 6 km/h, siendo este aumento más acusado en los meses de invierno, y el descenso en los meses de primavera y verano. Con todo esto observamos una tendencia menor que en la estación de Reus.

### A.E.3. Estudio regional

Ahora, para ver de una forma más intuitiva como se comporta el coeficiente  $\mu_1$  en el resto de estaciones, se dibujan los valores de  $\mu_1$  en cada mes para cada una de las estaciones de la Costa Nororiental y de la Costa Este. Realizadas las tablas 7 y 8 en el Anexo A.E.4 que muestran los valores de  $\mu_1$  a lo largo del año para las Costas Nororiental y Este, realizamos la figura 12.

A la vista de la figura 12, podemos observar que en la Costa Nororiental en todas las estaciones salvo en Barcelona Fabra se producen los valores máximos de  $\mu_1$  en los meses de otoño (con máximo en Octubre), y los mínimos se producen en Reus en Diciembre, mientras que en el resto se producen en verano. Cabe destacar que salvo Reus, las demás estaciones presentan similitudes en la gráfica y siguen el mismo comportamiento, siendo más acentuado en el área de Barcelona. Esto hace pensar con lo estudiado anteriormente que la tendencia es que el viento aumente a lo largo de los años en los meses de otoño y disminuya en los meses de verano.

En la Costa Este, los valores máximos de  $\mu_1$  se producen en los meses de otoño (con máximo en Noviembre) para las estaciones de Viveros y Castellón, y en Agosto para el Aeropuerto de Valencia. De la misma manera, los mínimos se producen en los meses de primavera (aunque tengamos mínimos en Diciembre) y en Febrero respectivamente. Es notable que en Viveros y Castellón los máximos y mínimos se obtengan con sólo un mes de diferencia. Cabe destacar que todas las estaciones presentan similitudes en la gráfica y siguen el mismo comportamiento, siendo más acentuado en Castellón. Esto hace pensar con lo estudiado anteriormente que la tendencia es que el viento aumente a lo largo de los años en otoño y disminuya en primavera.

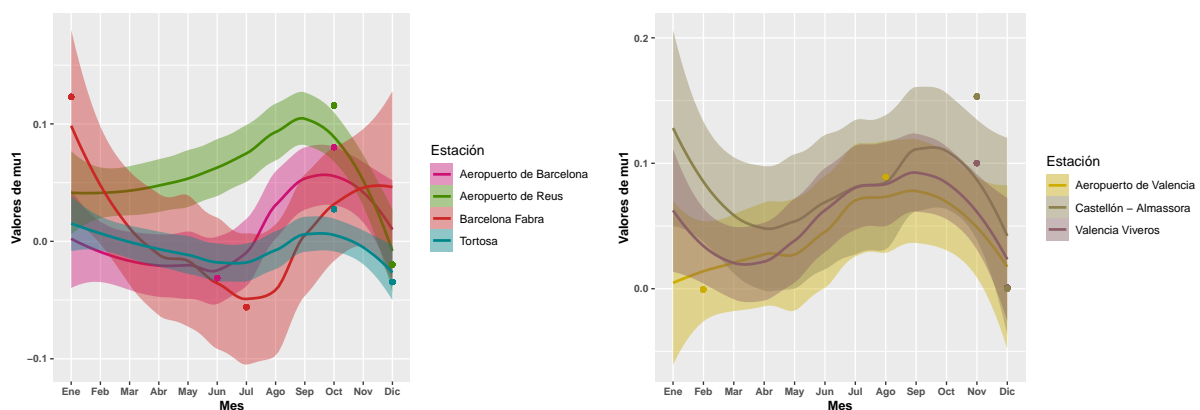


Figura 12: Valores de  $\mu_1$  a lo largo del año para la Costa Nororiental (izda.) y Costa Este (dcha.). Las curvas se ajustan con `geom_smooth` con el método `loess` con bandas de confianza del 95 %.

**A.E.4. Valores de  $\mu_1$** **Costa Nororiental**

En la tabla 7 tenemos los valores de  $\mu_1$  en las estaciones de la Costa Nororiental.

Mes	$\mu_1$			
	Tortosa	Reus	Barcelona	Fabra
Enero	0.02	0.04	0.01	0.12
Febrero	0.00	0.03	-0.02	0.01
Marzo	0.00	0.05	-0.03	0.00
Abril	0.00	0.05	0.00	-0.01
Mayo	-0.01	0.05	-0.02	-0.01
Junio	-0.02	0.05	-0.03	-0.03
Julio	-0.03	0.09	-0.02	-0.06
Agosto	0.00	0.07	0.03	-0.04
Septiembre	0.00	0.11	0.07	-0.01
Octubre	0.03	0.12	0.08	0.05
Noviembre	0.00	0.06	0.01	0.12
Diciembre	-0.03	-0.02	0.02	0.00

Tabla 7: Valores de  $\mu_1$  en las estaciones de la Costa Nororiental.

**Costa Este**

En la tabla 8 tenemos los valores de  $\mu_1$  en las estaciones de la Costa Este.

Mes	$\mu_1$		
	Valencia	Viveros	Castellón
Enero	0.00	0.07	0.14
Febrero	0.00	0.01	0.06
Marzo	0.07	0.03	0.08
Abril	0.02	0.01	0.02
Mayo	0.00	0.04	0.06
Junio	0.06	0.06	0.07
Julio	0.07	0.08	0.08
Agosto	0.09	0.09	0.09
Septiembre	0.04	0.07	0.08
Octubre	0.08	0.09	0.13
Noviembre	0.08	0.10	0.15
Diciembre	0.00	0.00	0.00

Tabla 8: Valores de  $\mu_1$  en las estaciones de la Costa Este.



## A.F. Modelización óptima

### A.F.1. Aeropuerto de Reus

#### Niveles de retorno para modelo inicial

En la tabla 9 tenemos unos valores de retorno que son más elevados en torno a los días de primavera elegidos, y menores en los días elegidos de invierno. También tenemos que tener en cuenta que en este modelo inicial, no tenemos interacción y que en todos los casos el cambio sea el mismo a lo largo del año tiene sentido.

Fecha	NDR-2		Cmb	NDR-10		Cmb	NDR-50		Cmb
	1961	2019	2	1961	2019	10	1961	2019	50
1 Ene	31.46	36.09	4.63	61.09	65.73	4.63	94.61	99.24	4.63
5 Mar	33.62	38.25	4.63	63.25	67.88	4.63	96.76	101.40	4.63
15 Abr	34.85	39.48	4.63	64.48	69.11	4.63	98.00	102.63	4.63
24 May	35.24	39.87	4.63	64.87	69.50	4.63	98.38	103.01	4.63
24 Jun	34.87	39.50	4.63	64.50	69.13	4.63	98.01	102.65	4.63
15 Ago	33.24	37.87	4.63	62.87	67.51	4.63	96.39	101.02	4.63
12 Oct	31.33	35.97	4.63	60.97	65.60	4.63	94.48	99.11	4.63
8 Dic	30.99	35.62	4.63	60.62	65.25	4.63	94.14	98.77	4.63

Tabla 9: Niveles de retorno a 2, 10 y 50 años para 1961 y 2019 junto con los cambios, es decir, la diferencia entre ambos valores, para las fechas seleccionadas.

#### Gráficos de diagnóstico estándar de los modelos inicial y finalista

En la figura 13, la primera gráfica podemos observar que las partes más superiores de las dos gráficas superiores son más problemáticas, que es justamente nuestra zona más interesada. También observamos que la gráfica de densidad obtenida no es satisfactoria. En la segunda gráfica los resultados son mucho más satisfactorios que los del modelo inicial. Pero para asegurarnos hasta qué punto es mejor, vamos a extraer la gráfica de los cuantiles empíricos para ver si mejora con respecto al modelo inicial. Para ello, se fija cuál de las dos tiene más valores por encima de 105 km/h, que es donde se desvían empíricos y estimados, y se considera el valor cuantil empírico asociado a ese valor, dibujando la vertical de ese valor en ambas gráficas. Cuanto más a la derecha esté esa vertical, mejor será el modelo.

### A.F.2. Aeropuerto de Almería

De forma análoga a Reus, consideramos el **modelo inicial**, formado por un **fevd** en el que intervienen los datos del Aeropuerto de Almería con indicador **X63250**, y **location.fun = ~ year + c1 + s1**, es decir, sólo interviene el parámetro de localización sin interacción, con unidades en km/h. De este modelo obtenemos un **valor AIC** de 184953,9.

Vemos los **niveles de retorno** estimados para 1961 y 2019 a 2, 10 y 50 años.

En la tabla 10 tenemos unos valores de retorno que son más elevados en torno a los días de primavera elegidos, y menores en los días elegidos de finales de verano y otoño, cercanos a los valores de los días de invierno. De nuevo, en este modelo inicial no tenemos interacción y el cambio es el mismo a lo largo del año.

En la tabla 11 se observan los valores AIC más bajos obtenidos en cada una de las tablas del Anexo A.F.4 para ver cuál es el menor, y nos quedaremos con ese modelo, siendo candidato de nuestro modelo final.

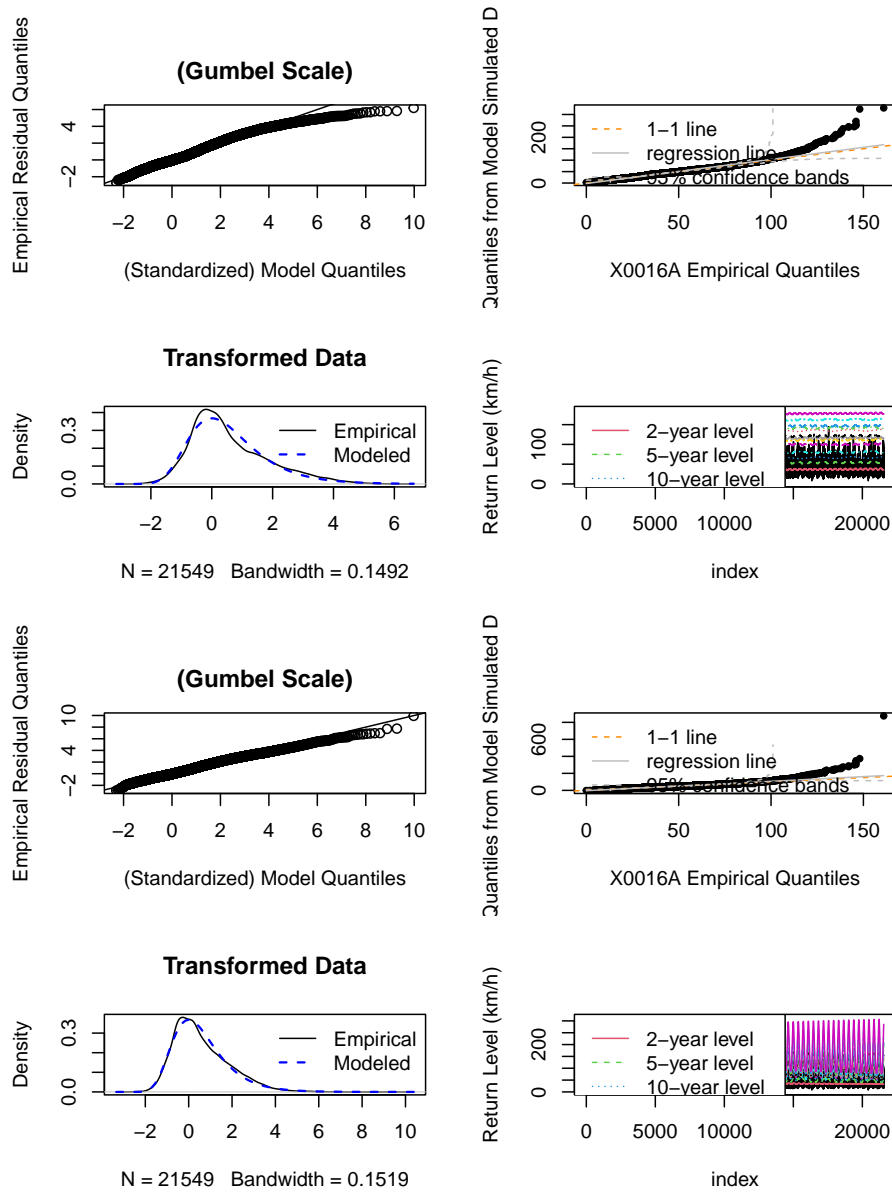


Figura 13: Modelo inicial arriba y modelo finalista abajo. El primer cuadro (arriba a la izquierda) muestra el gráfico cuantil-cuantil, en el segundo (arriba a la derecha) tenemos los cuantiles de una muestra extraída del GEV ajustado frente a los cuantiles de datos empíricos con bandas de confianza del 95 %, en el tercero (abajo a la izquierda) tenemos los gráficos de densidad de datos empíricos y el GEV ajustado, y por último (abajo a la derecha) tenemos el gráfico de nivel de retorno con los intervalos de confianza de aproximación normal puntual del 95 %.

De estos 3 modelos, vemos que el que tiene un menor valor AIC (184103,3) es el modelo con  $\text{location.fun} = \sim \text{year}*(c1 + s1)$ ,  $\text{scale.fun} = \sim \text{year} + c1 + s1$  y  $\text{shape.fun} = \sim \text{year}*(c1 + s1)$ , es decir, modelo en el que intervienen los tres parámetros, con el de localización  $\mu$  y forma  $\xi$  con expresión  $\text{year}*(c1 + s1)$ , y con el de escala  $\sigma$  con expresión  $\text{year} + c1 + s1$ . Aun así este modelo no equidista mucho del modelo intermedio, luego quizás debamos estudiar ambos modelos para ver cuál es mejor de los dos. Este va a ser nuestro modelo finalista, en el que viendo en R-studio [7] un resumen de sus datos obtenemos valores de  $\mu_1 = -3,488207 \cdot e^{-02}$ ,  $\sigma_1 = -9,691382 \cdot e^{-02}$  y  $\xi_1 = 4,071945 \cdot e^{-04}$ . La tendencia general, con valores similares a los de Reus, va descendiendo.

Vemos los gráficos de diagnóstico estándar de los modelos inicial y finalista para comparar.

Fecha	NDR-2		Cmb	NDR-10		Cmb	NDR-50		Cmb
	1961	2019	2	1961	2019	10	1961	2019	50
1 Ene	40.07	40.84	0.77	68.73	69.50	0.77	94.86	95.63	0.77
5 Mar	41.37	42.15	0.77	70.03	70.81	0.77	96.16	96.94	0.77
15 Abr	41.49	42.27	0.77	70.15	70.93	0.77	96.29	97.06	0.77
24 May	40.97	41.74	0.77	69.63	70.40	0.77	95.76	96.53	0.77
24 Jun	40.24	41.02	0.77	68.90	69.68	0.77	95.04	95.81	0.77
15 Ago	39.02	39.79	0.77	67.68	68.45	0.77	93.81	94.58	0.77
12 Oct	38.57	39.34	0.77	67.23	68.00	0.77	93.36	94.14	0.77
8 Dic	39.44	40.22	0.77	68.10	68.88	0.77	94.24	95.01	0.77

Tabla 10: Niveles de retorno a 2, 10 y 50 años para 1961 y 2019 junto con los cambios, para las fechas seleccionadas.

Valor	$2\sigma$	$2\mu + 2\sigma$	$2\mu + 1\sigma + 2\xi$
AIC	184448.9 (+0.18 %)	184137.4 (+0.01 %)	<b>184103.3</b> (100 %)
BIC	184512.8 (+0.15 %)	184241.1 (+0.005 %)	<b>184231.0</b> (100 %)

Tabla 11: Valores AIC más bajos obtenidos en cada una de las tablas de modelos para el Aeropuerto de Almería del Anexo A.F.4.

En la figura 14 las gráficas para el primer modelo nos salen bastante decentes y buenas, teniendo tendencia y estacionalidad sin interacción con la variable **year** y los armónicos **c1** y **s1**. Si lo comparamos con los modelos de Reus vemos que este modelo ya es mucho mejor. Para el segundo los resultados son peores que los del modelo inicial, hemos empeorado un poco las gráficas.

Por tanto, vamos a recordar que vimos en el exploratorio de Almería para encontrar una explicación:

Por un lado, vimos que el Aeropuerto de Almería tenía una marcada estacionalidad con valores medios altos en sus velocidades máximas diarias con respecto a otras estaciones cercanas a la zona como era el aeropuerto de Málaga. Además, su desviación típica era elevada prácticamente todo el año con variaciones no muy acusadas en verano e invierno. Por otro lado, vimos que era recomendable elegir el modelo sin distribución Gumbel, ya que los modelos arrojaban resultados similares pero el de distribución Gumbel tenía el parámetro de forma  $\xi = 0$  y era un modelo mas rígido. Además, vimos que había estacionalidad en los parámetros de localización y escala, luego a pesar de su rigidez, es conveniente estudiar los modelos con el parámetro de forma  $\xi = 0$  y la variable **year** con los armónicos **c1** y **s1**. Como hemos ya estudiado los valores AIC de estos modelos que nos interesan, y nos quedamos con los que tienen menor valor AIC, los modelos de la tabla final 11 son grandes candidatos, y de ellos el que tiene el segundo menor valor AIC (hay bastante diferencia con el que tiene mayor valor AIC de los tres) es un buen modelo para ser el candidato elegido para ver si es mejor que nuestro modelo final. Por tanto, vamos a estudiar ambos modelos para ver cuál es el mejor candidato.

Para ello, comenzamos consideramos el modelo sin el parámetro de forma de menor valor AIC obtenido en la tabla final como nuestro modelo candidato 1 ( $2\mu + 2\sigma$ ), con valores de  $\mu_1 = -0,030422622$ ,  $\sigma_1 = -0,094814523$ ; y el modelo sin el parámetro de forma de mayor valor AIC obtenido en la tabla final como nuestro modelo candidato 2 ( $2\sigma$ ) con valor de  $\sigma_1 = -0,084514417$ .

Para compararlos, veamos los gráficos de diagnóstico estándar de los ambos modelos junto con las gráficas de los cuantiles empíricos para ver los valores que tienen por encima de 90 y 100 km/h respectivamente.

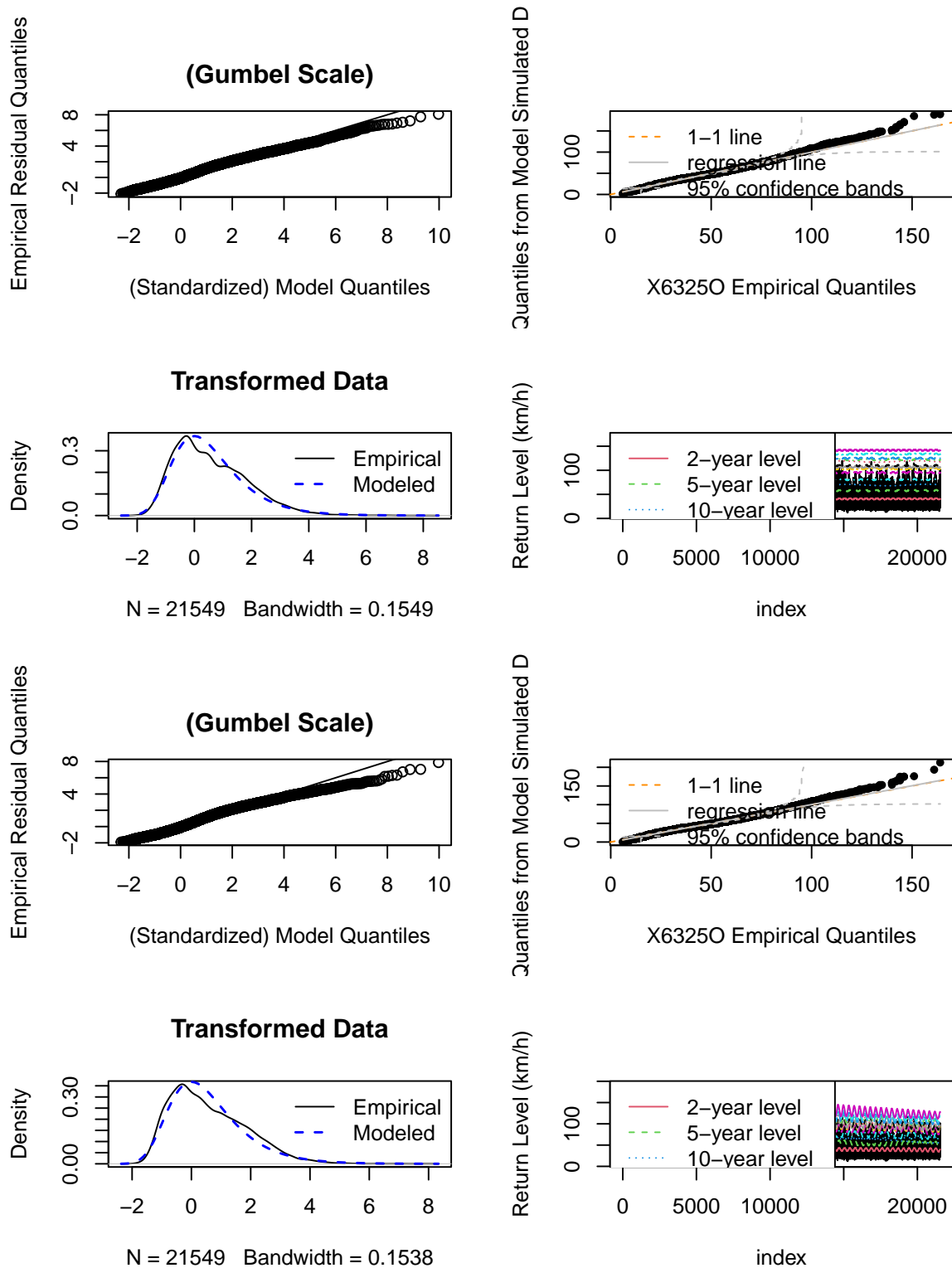


Figura 14: Modelo inicial (arriba) y modelo finalista (abajo). El primer cuadro (arriba a la izquierda) vemos el gráfico cuantil-cuantil, en el segundo (arriba a la derecha) tenemos los cuantiles de una muestra extraída del GEV ajustado frente a los cuantiles de datos empíricos con bandas de confianza del 95 %, en el tercero (abajo a la izquierda) tenemos los gráficos de densidad de datos empíricos y el GEV ajustado, y por último (abajo a la derecha) tenemos el gráfico de nivel de retorno con los intervalos de confianza de aproximación normal puntual del 95 %.

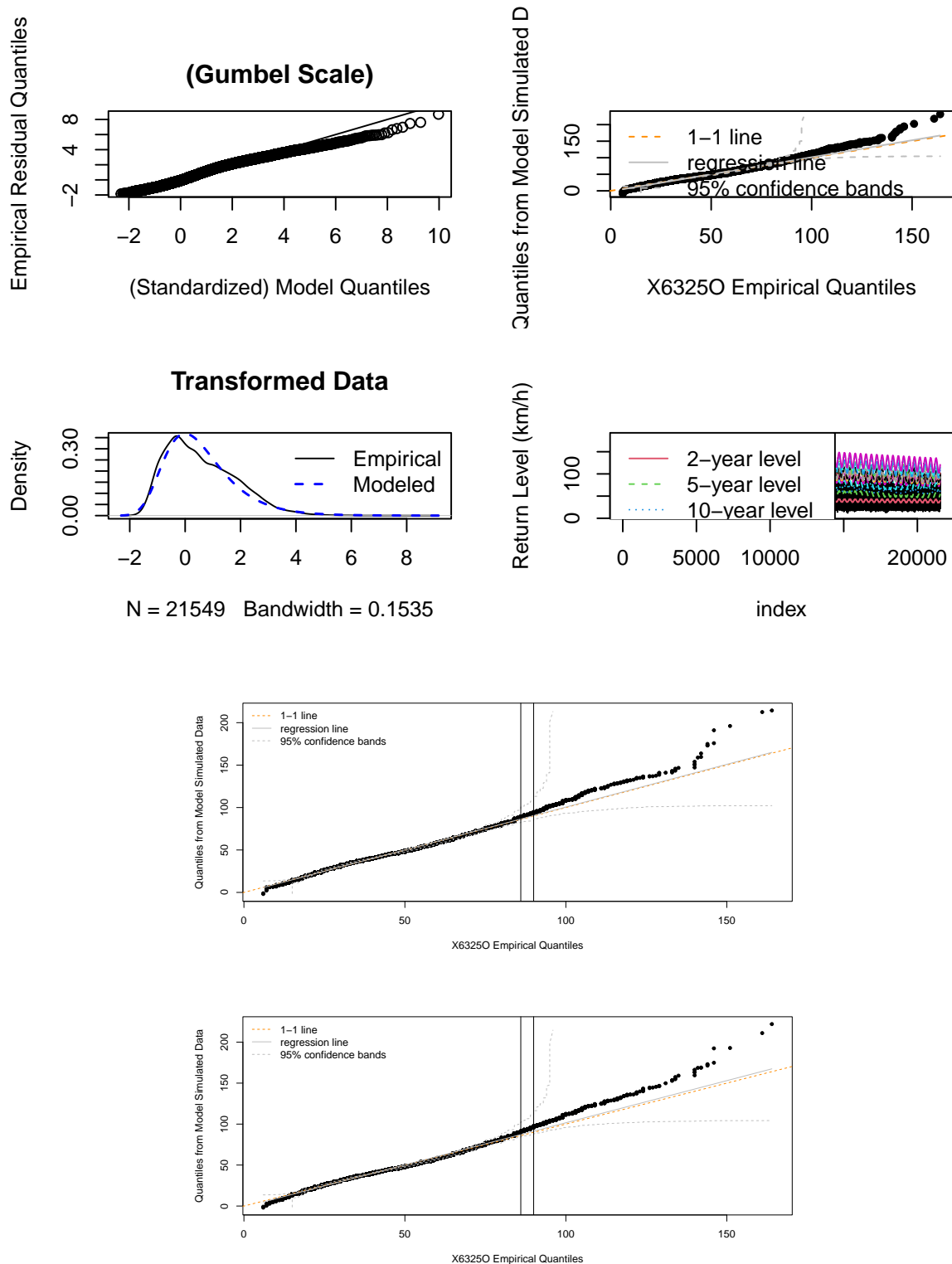


Figura 15: Modelo candidato 1. En cada una de las cuatro gráficas superiores, el primer cuadro (arriba a la izquierda) vemos el gráfico cuantil-cuantil, en el segundo (arriba a la derecha) tenemos los cuantiles de una muestra extraída del GEV ajustado frente a los cuantiles de datos empíricos con bandas de confianza del 95 %, en el tercero (abajo a la izquierda) tenemos los gráficos de densidad de datos empíricos y el GEV ajustado, y por último (abajo a la derecha) tenemos el gráfico de nivel de retorno con los intervalos de confianza de aproximación normal puntual del 95 %. En cada una de las inferiores, los cuantiles empíricos con dos verticales, una en los valores 90 y 100 respectivamente, y otra en el cuantil empírico asociado a ese valor.

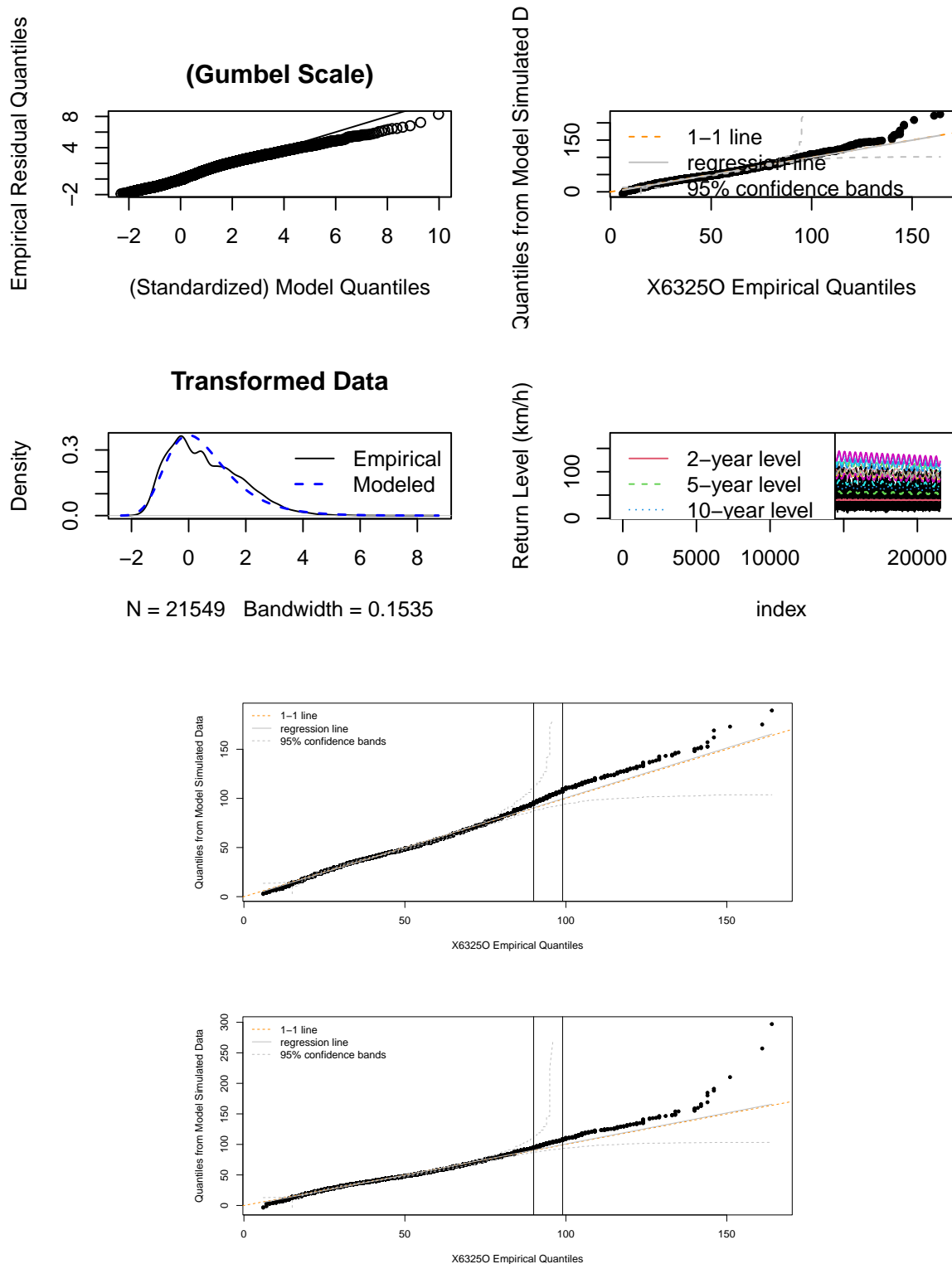


Figura 16: Modelo candidato 2. En cada una de las cuatro gráficas superiores, el primer cuadro (arriba a la izquierda) vemos el gráfico cuantil-cuantil, en el segundo (arriba a la derecha) tenemos los cuantiles de una muestra extraída del GEV ajustado frente a los cuantiles de datos empíricos con bandas de confianza del 95 %, en el tercero (abajo a la izquierda) tenemos los gráficos de densidad de datos empíricos y el GEV ajustado, y por último (abajo a la derecha) tenemos el gráfico de nivel de retorno con los intervalos de confianza de aproximación normal puntual del 95 %. En cada una de las inferiores, los cuantiles empíricos con dos verticales, una en los valores 90 y 100 respectivamente, y otra en el cuantil empírico asociado a ese valor.

Comparando las gráficas de ambos modelos en las figuras 15 y 16, vemos que prácticamente son iguales, por lo que nos vamos a quedar con el modelo que tiene menor AIC y que tiene el parámetro de forma  $\xi = 0$  con interacciones, ya que aparte de ser más fácil para trabajar, tiene en cuenta la estacionalidad por día del año y nos va a ayudar mejor a la hora de ver los niveles de retorno y sus cambios a lo largo del año. Por tanto, consideramos el modelo candidato 2 como nuestro **modelo final**.

Veamos los **niveles de retorno** estimados para 1961 y 2019 a 2, 10 y 50 años:

Fecha	NDR-2		Cmb	NDR-10		Cmb	NDR-50		Cmb
	1961	2019	2	1961	2019	10	1961	2019	50
1 Ene	42.40	39.96	-2.44	78.57	63.31	-15.25	111.20	84.38	-26.81
5 Mar	44.82	42.18	-2.64	83.09	68.68	-14.40	117.62	92.60	-25.02
15 Abr	45.10	41.60	-3.50	82.04	68.28	-13.76	115.37	92.35	-23.02
24 May	44.19	39.73	-4.46	78.41	64.99	-13.41	109.28	87.79	-21.50
24 Jun	42.89	37.86	-5.03	74.72	61.29	-13.43	103.44	82.42	-21.01
15 Ago	40.63	35.46	-5.18	69.88	55.87	-14.01	96.26	74.29	-21.97
12 Oct	39.75	35.60	-4.15	70.01	55.07	-14.94	97.32	72.64	-24.68
8 Dic	41.27	38.47	-2.81	75.60	60.25	-15.35	106.58	79.90	-26.67

Tabla 12: Niveles de retorno a 2, 10 y 50 años para 1961 y 2019 junto con los cambios, para las fechas seleccionadas.

En la tabla 12 tenemos unos valores de retorno que son más elevados en torno a la primera mitad del año a 2 años, siendo mayores en primavera y menores en verano y principios de otoño cuando los vemos a 10 y 50 años. También tenemos que tener en cuenta que en este modelo final, tenemos interacción y vemos que todos los cambios tienen tendencia a disminuir (aumentando estos valores negativos a medida que vemos los niveles de retorno a una cantidad de años mayor), siendo más acusados en los meses de invierno y menores en los de verano.

Seguidamente, de la misma manera que para Reus, vamos a ver una representación de los efectos hallados con el modelo final con la pendiente y expresiones de los parámetros (en este caso solamente  $\sigma$ ) que intervienen en el modelo en ambos años:

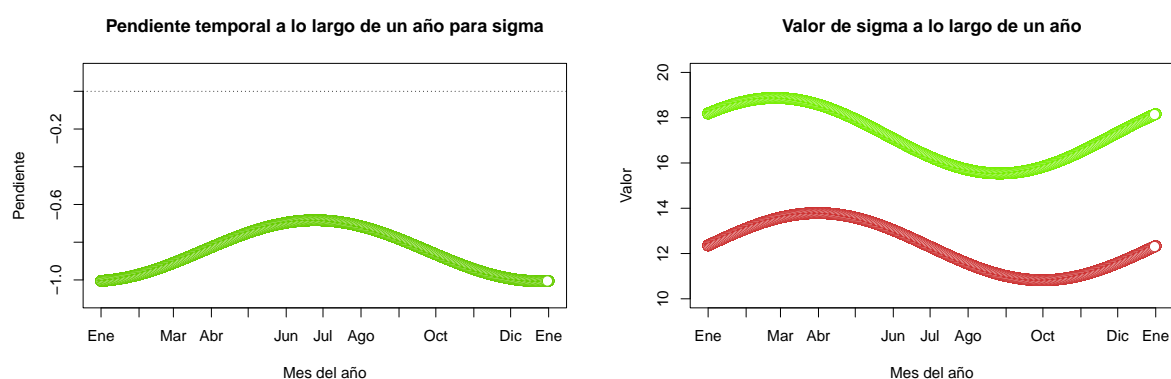


Figura 17: Izda., pendiente temporal en una década del parámetro  $\sigma$  con una línea horizontal en el valor 0. Dcha., expresión de  $\sigma$  en los años 1961 (verde) y 2019 (rojo). En ambas gráficas se expresa a lo largo de un año diariamente diferenciando en meses.

En la figura 17 observamos que las pendientes son negativas a lo largo de todo el año, con valores lejanos al 0, la escala cambia de manera muy similar a lo largo de todo el año, siendo menos negativas en los meses de verano. Para las expresiones de  $\sigma$ , vemos que los valores se han

modificado también, siendo menores en 2019 (aproximadamente una disminución de 6km/h), con el mismo comportamiento anual, pero con un retraso de unos dos meses entre ambos años, es decir, los máximos y mínimos en 2019 son unos dos meses más tarde que los máximos y mínimos de 1961.

### A.F.3. Barcelona Fabra

Consideramos el **modelo inicial**, formado por un **fevd** en el que intervienen los datos del Aeropuerto de Almería con indicador X0200E, y `location.fun = ~ year + c1 + s1`, es decir, sólo interviene el parámetro de localización sin interacción, con unidades en km/h. De este modelo obtenemos un **valor AIC** de 176260,2.

Vemos los **niveles de retorno** estimados para 1961 y 2019 a 2, 10 y 50 años:

Fecha	NDR-2		Cmb	NDR-10		Cmb	NDR-50		Cmb
	1961	2019	2	1961	2019	10	1961	2019	50
1 Ene	40.43	41.28	0.85	63.20	64.05	0.85	81.95	82.80	0.85
5 Mar	41.38	42.23	0.85	64.15	64.99	0.85	82.90	83.74	0.85
15 Abr	40.81	41.66	0.85	63.58	64.42	0.85	82.33	83.17	0.85
24 May	39.62	40.47	0.85	62.39	63.24	0.85	81.14	81.99	0.85
24 Jun	38.57	39.42	0.85	61.34	62.18	0.85	80.09	80.93	0.85
15 Ago	37.42	38.27	0.85	60.19	61.03	0.85	78.94	79.78	0.85
12 Oct	37.85	38.69	0.85	60.61	61.46	0.85	79.36	80.21	0.85
8 Dic	39.62	40.47	0.85	62.39	63.24	0.85	81.14	81.99	0.85

Tabla 13: Niveles de retorno a 2, 10 y 50 años para 1961 y 2019 junto con los cambios, para las fechas seleccionadas.

En la tabla 13 tenemos unos valores de retorno que son más elevados en torno a los días de primavera e invierno elegidos, y menores en los días elegidos de verano y otoño, aunque las diferencias no son tan notables. También tenemos que tener en cuenta que en este modelo inicial, no tenemos interacción y que en todos los casos el cambio sea el mismo a lo largo del año tiene sentido.

Veamos la tabla 14 con los valores AIC más bajos obtenidos en cada una de las tablas del Anexo A.F.4 para ver cuál es el menor, y nos quedaremos con ese modelo, siendo candidato de nuestro modelo final:

Valor	$2\sigma$	$2\mu + 2\sigma$	$2\mu + 2\sigma + 2\xi$
AIC	174896.7 (+0.36 %)	174318.6 (+0.03 %)	<b>174257.0</b> (100 %)
BIC	174960.5 (+0.32 %)	174422.3 (+0.01 %)	<b>174400.6</b> (100 %)

Tabla 14: Valores AIC más bajos obtenidos en cada una de las tablas de modelos para Barcelona Fabra del Anexo A.F.4.

De estos 3 modelos, vemos que el que tiene un menor valor AIC (174257) es el modelo con `location.fun = ~ year*(c1 + s1)`, `scale.fun = ~ year*(c1 + s1)` y `shape.fun = ~ year*(c1 + s1)`, es decir, modelo en el que intervienen los tres parámetros, de localización  $\mu$ , escala  $\sigma$  y forma  $\xi$  con expresión `year*(c1 + s1)`. Este va a ser nuestro modelo finalista, en el que viendo en R-studio [7] un resumen de sus datos obtenemos valores de  $\mu_1 = -1,343319 \cdot e^{-02}$ ,  $\sigma_1 = -1,193043 \cdot e^{-01}$  y  $\xi_1 = -1,537666 \cdot e^{-03}$ . Como en Reus, la tendencia va disminuyendo también.



Veamos ahora los gráficos de diagnóstico estándar de los modelos inicial y finalista para compararlos junto con las gráficas de los cuantiles empíricos de ambos modelos para ver los valores que tienen por encima de 90:

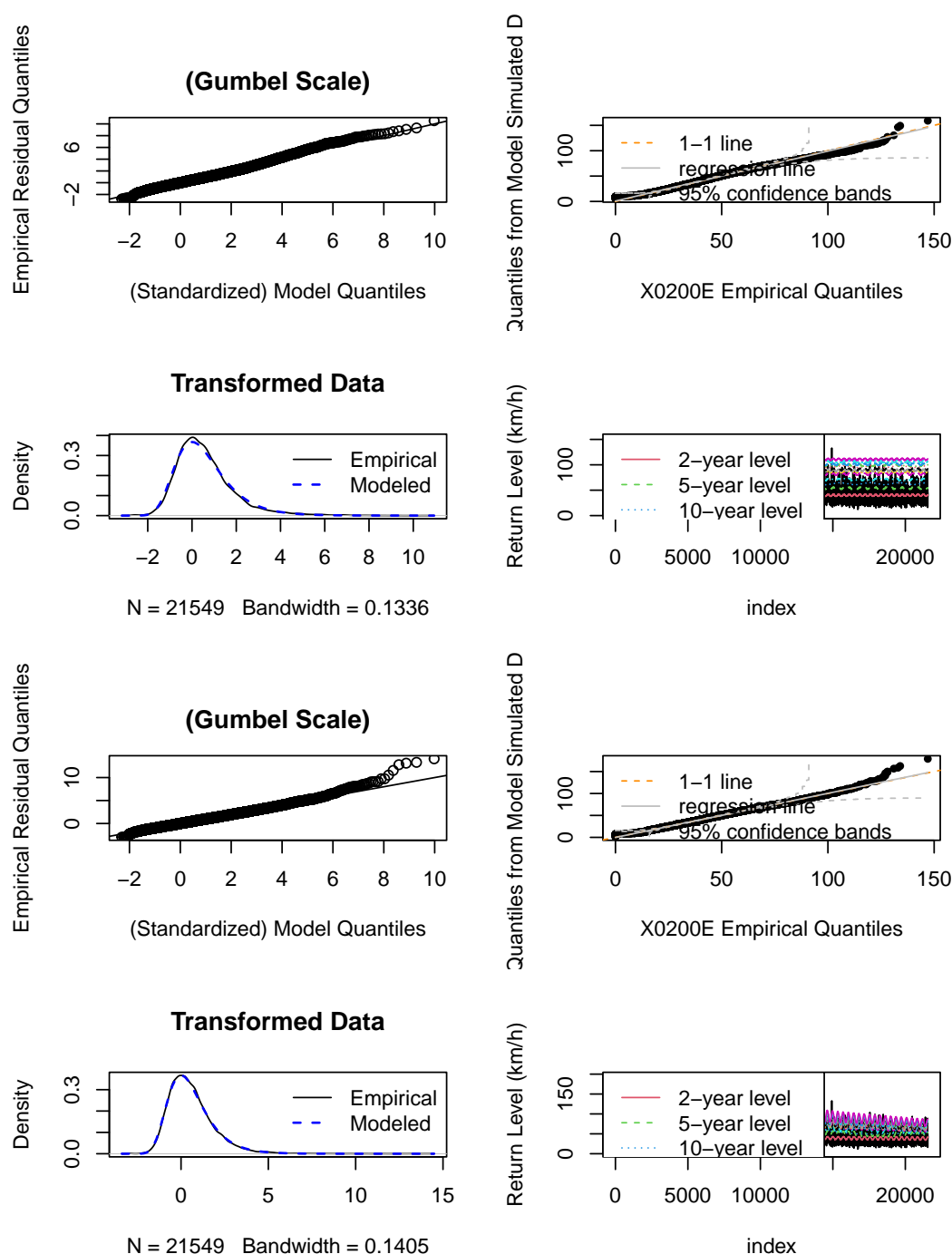


Figura 18: Modelo inicial arriba y modelo finalista abajo. En el primer cuadro (arriba a la izquierda) vemos el gráfico cuantil-cuantil, en el segundo (arriba a la derecha) tenemos los cuantiles de una muestra extraída del GEV ajustado frente a los cuantiles de datos empíricos con bandas de confianza del 95 %, en el tercero (abajo a la izquierda) tenemos los gráficos de densidad de datos empíricos y el GEV ajustado, y por último (abajo a la derecha) tenemos el gráfico de nivel de retorno con los intervalos de confianza de aproximación normal puntual del 95 %.

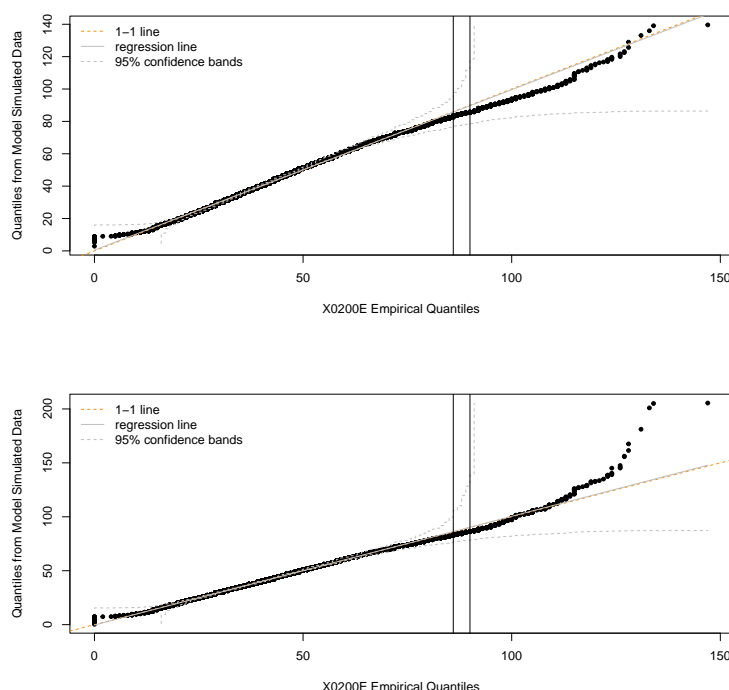


Figura 19: Cuantiles empíricos con dos verticales, una en el valor 90, y otra en el cuantil empírico asociado a ese valor.

Comparando ambas gráficas de las figuras 18 y 19, esta vez sí vemos que el modelo finalista es mucho mejor que el modelo inicial, los puntos se desvían menos. Por tanto, consideramos este último modelo finalista como nuestro **modelo final**.

Veamos los **niveles de retorno** estimados para 1961 y 2019 a 2, 10 y 50 años:

Fecha	NDR-2		Cmb	NDR-10		Cmb	NDR-50		Cmb
	1961	2019		1961	2019		1961	2019	
1 Ene	43.24	42.24	-0.99	81.34	59.77	-21.57	117.92	71.78	-46.14
5 Mar	43.93	41.60	-2.33	78.07	57.91	-20.16	110.33	69.07	-41.26
15 Abr	43.05	39.08	-3.97	71.52	53.64	-17.87	96.40	63.97	-32.44
24 May	41.62	36.34	-5.28	65.63	49.28	-16.35	84.87	58.87	-26.00
24 Jun	40.48	34.73	-5.75	62.76	46.92	-15.84	79.69	56.23	-23.47
15 Ago	39.45	34.37	-5.07	63.06	47.06	-16.00	80.72	56.90	-23.82
12 Oct	40.27	37.39	-2.87	70.34	52.58	-17.76	95.03	63.90	-31.13
8 Dic	42.38	41.22	-1.16	79.24	58.48	-20.76	113.43	70.54	-42.89

Tabla 15: Niveles de retorno a 2, 10 y 50 años para 1961 y 2019 junto con los cambios, para las fechas seleccionadas.

En la tabla 15 tenemos unos valores de retorno que son más elevados en los meses de invierno, y menores en los meses de verano. También tenemos que tener en cuenta que en este modelo final, tenemos interacción y vemos que todos los cambios tienden a disminuir, a 2 años son más acentuados en los meses de verano y menores en los de invierno, ocurriendo lo contrario a 10 y 50 años.

Como en Reus, se ve una representación de los efectos hallados con el modelo final con la pendiente y expresiones de los parámetros que intervienen en el modelo en ambos años.

Comenzamos con el parámetro  $\mu$ :

En la figura 20 observamos que las pendientes son negativas salvo en el periodo de invierno y finales de otoño, que son positivas. Para las expresiones de  $\mu$ , vemos que los valores son similares en ambos años con diferencias entre ambos no mayores de 4 km/h (en los meses de calor los valores de 2019 son superados por los de 1961). La gráfica es más acusada en 2019 con valores mínimos en verano y mayores en invierno, siendo menores en otoño y mayores en primavera en 1961.

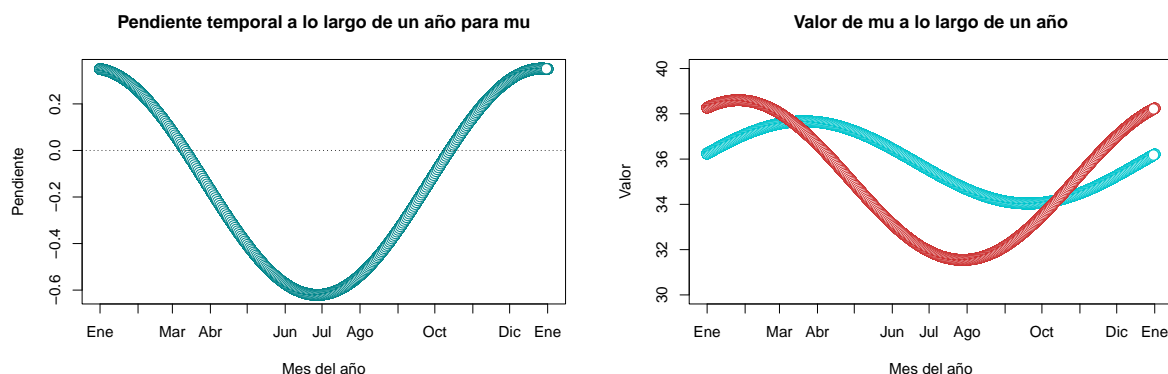


Figura 20: Izda., pendiente temporal en una década del parámetro  $\mu$  con una línea horizontal en el valor 0. Dcha., expresión de  $\mu$  en los años 1961 (azul) y 2019 (rojo). En ambas gráficas se expresa a lo largo de un año diariamente diferenciando en meses.

Continuamos con el parámetro  $\sigma$ :

En la figura 21 observamos que las pendientes son negativas a lo largo de todo el año, con valores lejanos al 0, la escala cambia de manera muy similar a lo largo de todo el año, siendo menos negativas entre los últimos meses de primavera y los primeros de verano. Para las expresiones de  $\sigma$ , vemos que los valores se han modificado también, siendo menores en 2019 (aproximadamente una disminución de 8km/h), pero con el mismo comportamiento anual.

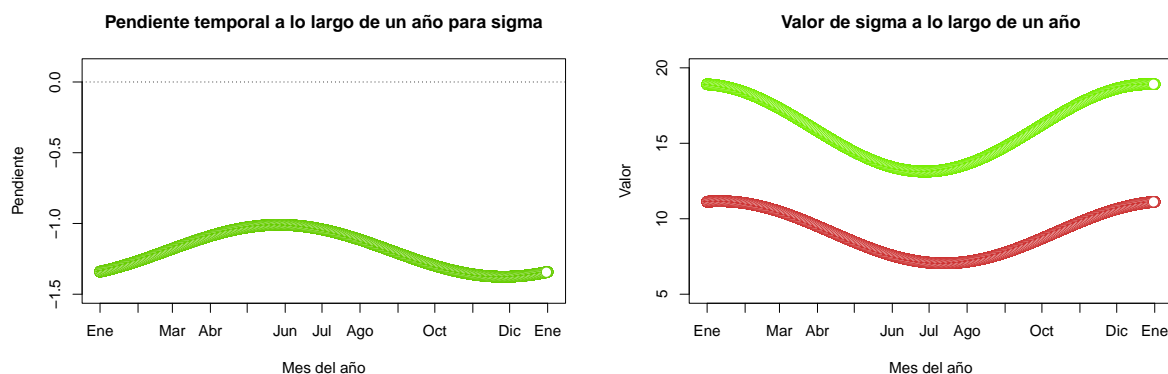


Figura 21: Izda., pendiente temporal en una década del parámetro  $\sigma$  con una línea horizontal en el valor 0. Dcha., expresión de  $\sigma$  en los años 1961 (verde) y 2019 (rojo). En ambas gráficas se expresa a lo largo de un año diariamente diferenciando en meses.

Finalizamos con el parámetro  $\xi$ :

En la figura 22 observamos que las pendientes son negativas en prácticamente todo el año salvo en los meses de verano, es decir, desde mediados de Junio hasta mediados de Septiembre. Aun así los valores son muy cercanos al 0, y por tanto la forma no cambia en prácticamente todo

el año. Para las expresiones de  $\xi$ , vemos que los valores no son muy diferentes numéricamente, pero sí que es cierto que los valores de 1961 son mayores que los de 2019 salvo en los meses de verano, aunque en este caso la diferencia entre ambos valores es mínima. Sin embargo si vemos que en 1961 los máximos se producen en invierno, los mínimos en verano, ocurriendo lo contrario en 2019, comportamiento opuesto.

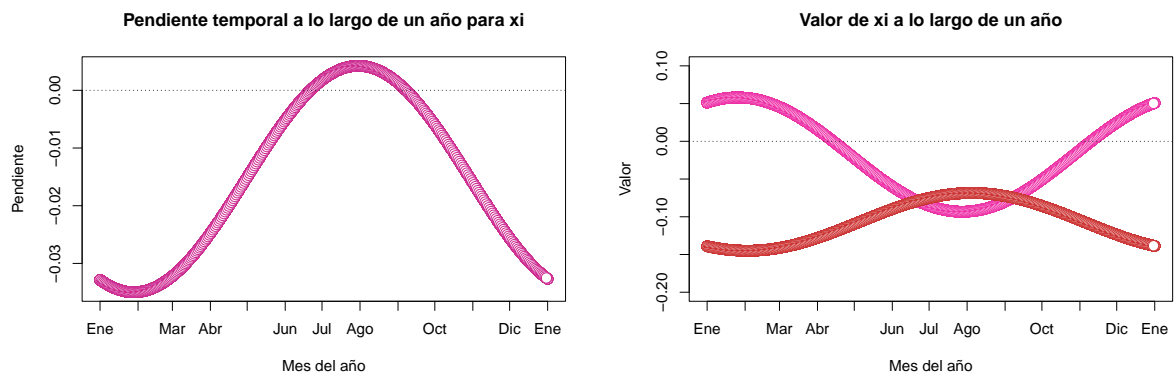


Figura 22: Izda., pendiente temporal en una década del parámetro  $\xi$ . Dcha., expresión de  $\xi$  en los años 1961 (morado) y 2019 (rojo). En ambas gráficas se expresa a lo largo de un año diariamente diferenciando en meses con una línea horizontal en el valor 0.

## A.F.4. Tablas de valores AIC

### Aeropuerto de Reus

A continuación vamos a ver la tabla de los valores AIC de los modelos en los que sólo interviene un parámetro:

Valor	$1\mu$	$2\mu$	$1\sigma$	$2\sigma$	$1\xi$	$2\xi$
AIC	182183.2	182161.4	179829.8	179821.2	182337.7	182313.8
BIC	182231.1	182225.2	179877.7	179885.1	182385.6	182377.6

Tabla 16: Valores AIC de los modelos formados con los datos del Aeropuerto de Reus en los que sólo interviene un parámetro.

De estos 6 modelos de la tabla 16, vemos que el que tiene un menor valor AIC (179821,2) es el modelo con `scale.fun = ~ year*(c1 + s1)`, es decir, modelo en el que interviene el parámetro de escala  $\sigma$  con expresión `year*(c1 + s1)`. Aquí vemos que la interacción no va a ser muy significativa, y no va a ser muy útil aunque sea el modelo que menor AIC hemos obtenido, al no haber mucha diferencia.

Ahora vamos a ver dos tablas en las que intervienen dos parámetros:

Valor	$1\mu + 1\sigma$	$1\mu + 2\sigma$	$1\mu + 1\xi$	$1\mu + 2\xi$	$2\mu + 1\sigma$	$2\mu + 2\sigma$
AIC	179618.6	179578.3	181719.3	181714.9	179499.2	179511.2
BIC	179690.4	179666.0	181791.1	181802.7	179587.0	179614.9

Valor	$2\mu + 1\xi$	$2\mu + 2\xi$	$1\sigma + 1\xi$	$1\sigma + 2\xi$	$2\sigma + 1\xi$	$2\sigma + 2\xi$
AIC	181722.8	181716.2	179536.0	179516.0	179530.3	179507.8
BIC	181810.6	181819.9	179607.8	179603.8	179618.0	179611.5

Tabla 17: Valores AIC de los modelos formados con los datos del Aeropuerto de Reus en los que intervienen dos parámetros.

De estos 12 modelos de la tabla 17, vemos que el que tiene un menor valor AIC (179499,2) es el modelo con `location.fun = ~ year*(c1 + s1)` y `scale.fun = ~ year + c1 + s1`, es decir, modelo en el que intervienen el parámetro de localización  $\mu$  con expresión `year*(c1 + s1)` y el de escala  $\sigma$  con expresión `year + c1 + s1`.

Por último, vemos dos tablas en las que intervienen los tres parámetros:

Valor	$1\mu + 1\sigma + 1\xi$	$1\mu + 1\sigma + 2\xi$	$1\mu + 2\sigma + 1\xi$	$1\mu + 2\sigma + 2\xi$
AIC	179323.3	179309.5	179314.7	179300.1
BIC	179419.1	179421.2	179426.4	179427.8

Valor	$2\mu + 1\sigma + 1\xi$	$2\mu + 1\sigma + 2\xi$	$2\mu + 2\sigma + 1\xi$	$2\mu + 2\sigma + 2\xi$
AIC	179232.6	179226.8	179231.7	179218.9
BIC	179344.3	179354.5	179359.3	179362.5

Tabla 18: Valores AIC de los modelos formados con los datos del Aeropuerto de Reus en los que interviene los tres parámetros.

De estos 8 modelos de la tabla 18, vemos que el que tiene un menor valor AIC (179218,9) es el modelo con `location.fun = ~ year*(c1 + s1)`, `scale.fun = ~ year*(c1 + s1)` y por último `shape.fun = ~ year*(c1 + s1)`, es decir, modelo en el que intervienen los tres parámetros, de localización  $\mu$ , escala  $\sigma$  y forma  $\xi$  con expresión `year*(c1 + s1)`.

### Aeropuerto de Almería

A continuación vamos a ver la tabla de los valores AIC de los modelos en los que sólo interviene un parámetro:

Valor	$1\mu$	$2\mu$	$1\sigma$	$2\sigma$	$1\xi$	$2\xi$
AIC	184953.9	185006.4	184461.3	184448.9	184976.5	184974.4
BIC	185001.8	185070.3	184509.1	184512.8	185024.4	185038.2

Tabla 19: Valores AIC de los modelos formados con los datos del Aeropuerto de Almería en los que sólo interviene un parámetro.

De estos 6 modelos de la tabla 19, vemos que el que tiene un menor valor AIC (184448,9) es el modelo con `scale.fun = ~ year*(c1 + s1)`, es decir, modelo en el que interviene el parámetro de escala  $\sigma$  con expresión `year*(c1 + s1)`.

Ahora vamos a ver dos tablas en las que intervienen dos parámetros:

Valor	$1\mu + 1\sigma$	$1\mu + 2\sigma$	$1\mu + 1\xi$	$1\mu + 2\xi$	$2\mu + 1\sigma$	$2\mu + 2\sigma$
AIC	184168.4	184151.6	184868.4	184866.6	184140.0	184137.4
BIC	184240.2	184239.4	184940.2	184954.4	184227.7	184241.1

Valor	$2\mu + 1\xi$	$2\mu + 2\xi$	$1\sigma + 1\xi$	$1\sigma + 2\xi$	$2\sigma + 1\xi$	$2\sigma + 2\xi$
AIC	184806.2	184968.2	184429.4	184433.0	184413.3	184413.8
BIC	184894.0	185071.9	184501.2	184520.7	184501.0	184517.5

Tabla 20: Valores AIC de los modelos formados con los datos del Aeropuerto de Almería en los que intervienen dos parámetros.

De estos 12 modelos de la tabla 20, vemos que el que tiene un menor valor AIC (184137,4) es el modelo con `location.fun = ~ year*(c1 + s1)` y `scale.fun = ~ year*(c1 + s1)`, es decir, modelo en el que intervienen los parámetros de localización  $\mu$  y escala  $\sigma$  con expresión `year*(c1 + s1)`.

Por último, vemos dos tablas en las que intervienen los tres parámetros:

Valor	$1\mu + 1\sigma + 1\xi$	$1\mu + 1\sigma + 2\xi$	$1\mu + 2\sigma + 1\xi$	$1\mu + 2\sigma + 2\xi$
AIC	184141.3	184142.7	184123.5	184852.8
BIC	184237.0	184254.4	184235.1	184980.5

Valor	$2\mu + 1\sigma + 1\xi$	$2\mu + 1\sigma + 2\xi$	$2\mu + 2\sigma + 1\xi$	$2\mu + 2\sigma + 2\xi$
AIC	184107.1	184103.3	184106.8	184607.7
BIC	184218.8	184231.0	184234.5	184751.3

Tabla 21: Valores AIC de los modelos formados con los datos del Aeropuerto de Almería en los que intervienen los tres parámetros.

De estos 8 modelos de la tabla 21, vemos que el que tiene un menor valor AIC (184103,3) es el modelo con `location.fun = ~ year*(c1 + s1)`, `scale.fun = ~ year + c1 + s1` y por último `shape.fun = ~ year*(c1 + s1)`, es decir, modelo en el que intervienen los tres parámetros, con el de localización  $\mu$  y forma  $\xi$  con expresión `year*(c1 + s1)`, y con el de escala  $\sigma$  con expresión `year + c1 + s1`.

### Barcelona Fabra

A continuación vamos a ver la tabla de los valores AIC de los modelos en los que sólo interviene un parámetro:

Valor	$1\mu$	$2\mu$	$1\sigma$	$2\sigma$	$1\xi$	$2\xi$
AIC	176260.1	176166.5	174906.4	174896.7	176229.7	176202.8
BIC	176308.0	176230.3	174954.2	174960.5	176277.6	176266.6

Tabla 22: Valores AIC de los modelos formados con los datos de Barcelona Fabra en los que sólo interviene un parámetro.

De estos 6 modelos de la tabla 22, vemos que el que tiene un menor valor AIC (174896,7) es el modelo con `scale.fun = ~ year*(c1 + s1)`, es decir, modelo en el que interviene el parámetro de escala  $\sigma$  con expresión `year*(c1 + s1)`.

Ahora vamos a ver dos tablas en las que intervienen dos parámetros:

Valor	$1\mu + 1\sigma$	$1\mu + 2\sigma$	$1\mu + 1\xi$	$1\mu + 2\xi$	$2\mu + 1\sigma$	$2\mu + 2\sigma$
AIC	174405.5	174351.7	176002.8	175980.8	174343.4	174318.6
BIC	174477.3	174439.5	176074.6	176068.6	174431.2	174422.3

Valor	$2\mu + 1\xi$	$2\mu + 2\xi$	$1\sigma + 1\xi$	$1\sigma + 2\xi$	$2\sigma + 1\xi$	$2\sigma + 2\xi$
AIC	175804.9	175761.8	174796.4	174762.8	174784.0	174760.1
BIC	175892.6	175865.5	174868.2	174850.6	174871.7	174863.9

Tabla 23: Valores AIC de los modelos formados con los datos de Barcelona Fabra en los que intervienen dos parámetros.

De estos 12 modelos de la tabla 23, vemos que el que tiene un menor valor AIC (174318,6) es el modelo con `location.fun = ~ year*(c1 + s1)` y `scale.fun = ~ year*(c1 + s1)`, es decir, modelo en el que intervienen los parámetros de localización  $\mu$  y escala  $\sigma$  con expresión `year*(c1 + s1)`.

Por último, vemos dos tablas en las que intervienen los tres parámetros:

Valor	$1\mu + 1\sigma + 1\xi$	$1\mu + 1\sigma + 2\xi$	$1\mu + 2\sigma + 1\xi$	$1\mu + 2\sigma + 2\xi$
AIC	174364.2	174342.6	174348.8	175699.3
BIC	174459.9	174454.3	174460.4	175827.0

Valor	$2\mu + 1\sigma + 1\xi$	$2\mu + 1\sigma + 2\xi$	$2\mu + 2\sigma + 1\xi$	$2\mu + 2\sigma + 2\xi$
AIC	174342.9	174264.7	174291.9	174257.0
BIC	174454.6	174392.3	174419.6	174400.6

Tabla 24: Valores AIC de los modelos formados con los datos de Barcelona Fabra en los que intervienen los tres parámetros.

De estos 8 modelos de la tabla 24, vemos que el que tiene un menor valor AIC (174257) es el modelo con `location.fun = ~ year*(c1 + s1)`, `scale.fun = ~ year*(c1 + s1)` y por último `shape.fun = ~ year*(c1 + s1)`, es decir, modelo en el que intervienen los tres parámetros, de localización  $\mu$ , escala  $\sigma$  y forma  $\xi$  con expresión `year*(c1 + s1)`.



## A.G. Conclusiones

### A.G.1. Tablas informativas de cuantiles

#### Aeropuerto de Reus

Fecha	1961		2019		Cambio absoluto		Cambio relativo	
	q50	q90	q50	q90	q50	q90	q50	q90
1 de Enero	41.42	89.34	33.05	59.98	-8.37	-29.36	-0.20	-0.33
5 de Marzo	43.18	84.50	34.27	56.12	-8.91	-28.38	-0.21	-0.34
15 de Abril	41.35	74.09	34.62	50.13	-6.73	-23.96	-0.16	-0.32
24 de Mayo	38.22	64.31	34.43	45.11	-3.79	-19.20	-0.10	-0.30
24 de Junio	35.65	58.80	33.94	42.50	-1.71	-16.30	-0.05	-0.28
15 de Agosto	33.17	57.27	32.85	42.13	-0.32	-15.14	-0.01	-0.26
12 de Octubre	34.80	68.76	32.13	48.76	-2.67	-20.00	-0.08	-0.29
8 de Diciembre	39.48	85.26	32.58	57.98	-6.90	-27.28	-0.17	-0.32

Tabla 25: Cuantiles 50 y 90 estimados en 1961 y 2019, los cambios absolutos y relativos de éstos.

#### Aeropuerto de Almería

Fecha	1961		2019		Cambio absoluto		Cambio relativo	
	q50	q90	q50	q90	q50	q90	q50	q90
1 de Enero	41.40	76.53	39.26	63.13	-2.14	-13.40	-0.05	-0.18
5 de Marzo	41.65	78.07	39.73	66.07	-1.92	-12.00	-0.05	-0.15
15 de Abril	41.45	76.82	39.77	66.30	-1.68	-10.52	-0.04	-0.14
24 de Mayo	41.08	74.49	39.57	65.06	-1.51	-9.43	-0.04	-0.13
24 de Junio	40.76	72.52	39.30	63.40	-1.46	-9.12	-0.04	-0.13
15 de Agosto	40.44	70.52	38.86	60.62	-1.58	-9.90	-0.04	-0.14
12 de Octubre	40.61	71.58	38.70	59.66	-1.91	-11.92	-0.05	-0.17
8 de Diciembre	41.16	75.03	39.03	61.70	-2.13	-13.34	-0.05	-0.18

Tabla 26: Cuantiles 50 y 90 en 1961 y 2019 y los cambios absolutos y relativos de éstos.

**Barcelona Fabra**

Fecha	1961		2019		Cambio absoluto		Cambio relativo	
	q50	q90	q50	q90	q50	q90	q50	q90
1 de Enero	43.24	81.34	42.39	64.82	-0.85	-16.52	-0.02	-0.20
5 de Marzo	43.93	78.07	41.72	62.38	-2.21	-15.69	-0.05	-0.20
15 de Abril	43.05	71.52	39.16	56.11	-3.89	-15.40	-0.09	-0.22
24 de Mayo	41.62	65.63	36.37	50.05	-5.26	-15.58	-0.13	-0.24
24 de Junio	40.48	62.76	34.73	46.87	-5.75	-15.89	-0.14	-0.25
15 de Agosto	39.45	63.06	34.36	46.72	-5.08	-16.34	-0.13	-0.26
12 de Octubre	40.27	70.34	37.42	53.72	-2.84	-16.62	-0.07	-0.24
8 de Diciembre	42.38	79.24	41.34	62.55	-1.04	-16.69	-0.02	-0.21

Tabla 27: Cuantiles 50 y 90 en 1961 y 2019 y los cambios absolutos y relativos de éstos.

**A.G.2. Gráficas de los cambios absolutos y relativos en las tres estaciones**

Finalmente, vamos a ver las gráficas de los cambios absolutos y relativos asociados a un día concreto del año, el 15 de Abril, para ver las diferencias entre las tres estaciones, dibujando por grupos mediante un ggplot con el método `loess`.

En la figura 23, vemos que las disminuciones más acentuadas se producen en el cuantil 90, y en mayor magnitud en el Aeropuerto de Reus, estación con más movimiento climático. Aun así, vemos claramente una reducción generalizada en ambos cambios.

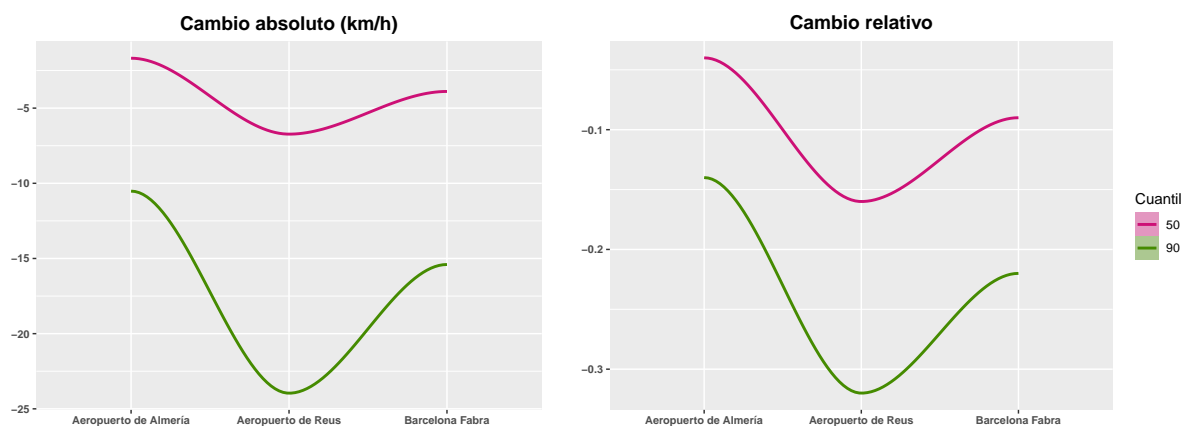


Figura 23: Cambio absoluto izda y cambio relativo dcha.