

Propiedades probabilísticas y análisis de la ocurrencia de récords



Juan Marcén Gutiérrez
Trabajo de fin de grado de Matemáticas
Universidad de Zaragoza

Directora del trabajo: Ana C. Cebrián Guajardo
9 de julio de 2024

Resumen

El objetivo de este trabajo se centra en desarrollar herramientas estadísticas para detectar desviaciones de un comportamiento estacionario en las colas de una distribución, en particular en la ocurrencia de récords. Estas herramientas están basadas en las propiedades de la distribución de la ocurrencia de récords en series de variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.).

Para lograr este objetivo, primero, se tienen que revisar los principales resultados sobre la distribución de récords. Se trabajará con el *Modelo Clásico de Récords (MCR)*, donde $(X_t)_{t \geq 1}$ es una serie de v.a.i.i.d. En el Capítulo 1, se muestran definiciones y propiedades de dichas series. Una observación X_i se dice *récord superior* si su valor excede a los anteriores. Análogamente se definen los *récords inferiores*. Se definen la serie de variables aleatorias *indicadoras de récord* $(I_t)_{t \geq 1}$ como $I_t = 1$ si X_t es un récord e $I_t = 0$ si no. El *número de récords* $N_t)_{t \geq 1}$ indicará el número de récords producidos hasta el instante t . Este se define como $N_t = \sum_{i=1}^t I_i$. A continuación se tratan propiedades de gran relevancia para el trabajo sobre las variables I_t y N_t bajo el MCR. Para ambos tipos de variables se demuestra como su distribución probabilística no depende de la serie original (X_t) . Dicha propiedad resulta de gran importancia para crear herramientas estadísticas no paramétricas, las cuales se pueden aplicar a cualquier tipo de series. Las variables indicadoras de récord tienen la siguiente distribución.

$$P(I_t = 1) = \frac{1}{t}, \quad t = 1, 2, \dots$$

Para el caso de los números de récords N_t , se obtiene una distribución más compleja de analizar. Por ello se usa el Teorema Central del Límite de Lyapunov para obtener una distribución asintótica.

$$\frac{N_t - \mu_t}{\sigma_t} \xrightarrow{D} N(0, 1),$$

donde

$$\mu_t = E(N_t) = \sum_{j=1}^t \frac{1}{j} \quad \text{y} \quad \sigma_t^2 = \text{Var}(N_t) = \sum_{j=1}^t \left(\frac{1}{j} - \frac{1}{j^2} \right).$$

En el Capítulo 2 de este trabajo se desarrollan herramientas estadísticas para detectar desviaciones de un comportamiento estacionario en las colas de la distribución de la serie de ocurrencia de récords de una serie (X_t) bajo la hipótesis nula de que es una serie de v.a.i.i.d. ($P(I_t = 1) = 1/t$). La hipótesis alternativa se resume a que la probabilidad de ocurrencia de récord sea distinta a $1/t$. En particular, si la hipótesis alternativa es que haya efectos de una tendencia creciente en la serie de ocurrencia de récords, esta sería $p_{tm} > 1/t$. Para detectar una tendencia decreciente se utilizaría $p_{tm} < 1/t$. Para poder aplicar dichas herramientas, se supondrán M series independientes de longitud T . La obtención de dichas series se explica en el Capítulo 3. Una vez se han definido las hipótesis, se presentan diferentes estadísticos para realizar los respectivos contrastes de hipótesis. Por un lado se presentan contrastes basados en un tipo de récords. Primero se muestran contrastes basados en el número de récords N_t . Diersen y Trenkler (1996) dedujeron un estadístico muy potente,

$$\mathcal{N}^w = \sum_{t=1}^T \sum_{m=1}^M w_t I_t = \sum_{t=1}^T w_t S_t,$$

donde w_t es una función creciente en t . Los autores recomendaron $w_t = t - 1$. Se crea un contraste, donde el p -valor es

$$P\left(Z > \frac{\mathcal{N}_0^w - e_{\mathcal{N}^w} - 0,5}{\sqrt{v_{\mathcal{N}^w}}}\right),$$

donde $Z \sim N(0, 1)$, \mathcal{N}_0^w es el valor observado de \mathcal{N}^w , 0,5 es un corrector de continuidad, $e_{\mathcal{N}^w}$ es la esperanza del estadístico y $v_{\mathcal{N}^w}$ la varianza.

También se propone un contraste en el que el estadístico esta basado en las variables indicadoras de récords I_t , en particular de su función de verosimilitud. El estadístico a estudiar se construye con la suma de los vectores score bajo la hipótesis nula.

$$\mathcal{S} = \frac{\sum_{t=2}^T t(tS_t - M)/(t-1)}{\sqrt{M \sum_{t=2}^T t^2/(t-1)}}.$$

Este estadístico puede ser considerado un caso particular de \mathcal{N}^w , y además, es asintóticamente normal cuando $M \rightarrow \infty$. Contrastes basados en este estadístico resultan de fácil aplicación debido a que su esperanza y varianza bajo la hipótesis nula es igual a 0 y 1 respectivamente. Dicho contraste es el contraste insesgado más potente, el cual resulta equivalente a utilizar pesos $w_1 = 0, w_t = t^2/(t-1), t = 2, \dots, T$ en el estadístico \mathcal{N}^w .

Por otro lado, debido al problema presente de escasez de datos en las colas de dichas series, se presentan unos contrastes más potentes. Dichos contrastes utilizan la información de cuatro tipos de récords. Las series de récords superiores (U) e inferiores (L) hacia delante (F) y hacia atrás (B). La serie de récords hacia atrás se define como la serie original pero empezando desde la última observación, i.e. X_T, X_{T-1}, \dots, X_1 . Para cada tipo de serie, se calcula uno de los estadísticos anteriores. En este trabajo, se muestra para \mathcal{N}^w . Para unificar el comportamiento de los cuatro estadísticos, se considerarán combinaciones lineales de \mathcal{N}^w . Para estudiar el efecto de una tendencia positiva en las colas superiores tenemos

$$U = \mathcal{N}^{wFU} - \mathcal{N}^{wBU} = \sum_{m=1}^M \sum_{t=1}^T w_t (I_{tm}^{FU} - I_{tm}^{BU}).$$

Para colas inferiores

$$L = \mathcal{N}^{wBL} - \mathcal{N}^{wFL} = \sum_{m=1}^M \sum_{t=1}^T w_t (I_{tm}^{BL} - I_{tm}^{FL}).$$

Para ambas colas

$$D = \mathcal{N}^{wFU} - \mathcal{N}^{wFL} - \mathcal{N}^{wBU} + \mathcal{N}^{wBL} = \sum_{m=1}^M \sum_{t=1}^T w_t (I_{tm}^{FU} - I_{tm}^{FL} - I_{tm}^{BU} + I_{tm}^{BL}).$$

Por último, se propone una modificación de los contrastes anteriores para poder aplicarlos sobre M series dependientes.

En el Capítulo 3 se realiza un análisis estadístico de datos con las herramientas del Capítulo 2 y herramientas gráficas implementadas en la librería de R *RecordTest* con el fin de detectar los efectos del calentamiento global. Se estudiarán las series de temperaturas máximas y mínimas diarias en las localidades de Madrid, Daroca, Castellón, Tortosa, Izaña, Huelva y Cáceres, utilizando como ejemplo a Madrid. Primero, se explican los procesos de transformación de datos para la obtención de M series independientes sin componente estacional de longitud T . Tras la transformación de datos, se realiza un análisis, el cual consiste en estudiar el número de récords en las colas de la serie de temperaturas máximas y mínimas diarias de Madrid. También, al final del capítulo, se presenta un resumen para el resto de localidades. Los resultados obtenidos, tanto gráfica y numéricamente, en los contrastes realizados para todas localidades salvo Cáceres, sugieren la existencia de una tendencia positiva en las colas de dichas series, lo cual invita a la existencia de cambio climático.

Abstract

The objective of this work focuses on the development of statistical tools to detect non-stationary behaviours on the tails of a distribution, particularly in the record occurrence series. These tools are based on the properties of the distribution of the occurrence of records in independent and identically distributed (i.i.d.) series.

To achieve this goal, the main results of the distribution of records need to be reviewed. The *Classical Model of Records (CMR)* will be worked with, where $(X_t)_{t \geq 1}$ is a series of i.i.d. variables. In Chapter 1, definitions and properties of such series are showed. An observation X_i is said to be an *upper record* if its value exceeds the previous ones. Analogously, *lower records* are defined. The series of *record indicator* variables $(I_t)_{t \geq 1}$ are defined, where $I_t = 1$ if X_t is a record and $I_t = 0$ otherwise. The *number of records* $(N_t)_{t \geq 1}$ indicates the number of records produced up to instant t . These variables are defined as $N_t = \sum_{i=1}^t I_i$. Next, under the *CMR*, properties of great relevance for our work for the variables I_t and N_t are treated. For both variables, the independence of their distribution from the original series (X_t) is proved. Such property is of great importance in order to create nonparametric statistical tools, which can be applied to any type of series. Record indicator variables have the following distribution.

$$P(I_t = 1) = \frac{1}{t}, \quad t = 1, 2, \dots$$

The number of records N_t variables have a more complex distribution. Therefore, the Lyapunov's Central Limit Theorem is used in order to obtain an assymptotic distribution.

$$\frac{N_t - \mu_t}{\sigma_t} \xrightarrow{D} N(0, 1),$$

where

$$\mu_t = E(N_t) = \sum_{j=1}^t \frac{1}{j} \quad \text{and} \quad \sigma_t^2 = Var(N_t) = \sum_{j=1}^t \left(\frac{1}{j} - \frac{1}{j^2} \right).$$

In Chapter 2 several statistical tools to study a non-stationary behaviour in the tails of the distribution of the record occurrence series are developed. These tools are under the null hypothesis of this series being an i.i.d. series ($P(I_t = 1) = 1/t$). The alternative hypothesis will be that the probability of record occurrence is different from $1/t$. In the case of studying the effects of an increasing trend, the alternative will be that the probability of record is greater than $1/t$. On the other hand, the analysis of effects of a decreasing trend will suppose that the alternative will be lower than $1/t$. In order to apply such tools, it is supposed to have M independent series of length T available. The derivation of these series is explained in Chapter 3. After the hypothesis are defined, different statistics are presented in order to perform the respective hypothesis tests. First, tests based on the number of records N_t are shown. Diersen and Trenkler (1996) derived a very powerful statistic,

$$\mathcal{N}^w = \sum_{t=1}^T \sum_{m=1}^M w_t I_t = \sum_{t=1}^T w_t S_t,$$

where w_t is an increasing function in t . The authors recommended the use of $w_t = t - 1$. This statistic is asymptotically normal when $M \rightarrow \infty$. A test is created, where the p -value is

$$P\left(Z > \frac{\mathcal{N}_0^w - e_{\mathcal{N}^w} - 0,5}{\sqrt{v_{\mathcal{N}^w}}}\right),$$

where $Z \sim N(0, 1)$, \mathcal{N}_0^w is the observed value of the statistic, $e_{\mathcal{N}^w}$ its expectation and $v_{\mathcal{N}^w}$ its variance. Also a continuity correction is applied.

A test based on the record indicator variables I_t is also shown. The statistic to be studied is constructed by the sum of the score vectors under the null hypothesis (i.i.d. series).

$$\mathcal{S} = \frac{\sum_{t=2}^T t(tS_t - M)/(t-1)}{\sqrt{M \sum_{t=2}^T t^2/(t-1)}}.$$

This statistic can be considered as a particular case of \mathcal{N}^w and it is asymptotically normal when $M \rightarrow \infty$. Testing with such statistic is easier since, under the null, its expectation and variance are equal to 0 and 1, respectively. This test is the most powerful unbiased test, which is equivalent to use weights $w_1 = 0$, $w_t = t^2/(t-1)$, $t = 2, \dots, T$ in \mathcal{N}^w .

On the other hand, due to the present problem of scarcity of data in the tails of the series, more powerful tests have been developed. These tests join the information from four types of records, upper (U) and lower (L) records of forward (F) and backwards (B) series. The backward series are the series obtained when the order of the terms are reversed so we start by observing the last term, i.e. X_T, X_{T-1}, \dots, X_1 . For each type of series, one of the previous statistics will be computed. In our case, we will work with the statistic \mathcal{N}^w . To unify the behaviour of the four statistics, we will consider linear combinations of \mathcal{N}^w . To study the effect of a positive trend in the upper tails we will consider

$$U = \mathcal{N}^{wFU} - \mathcal{N}^{wBU} = \sum_{m=1}^M \sum_{t=1}^T w_t (I_{tm}^{FU} - I_{tm}^{BU}).$$

To study this effect in the lower tails we have

$$L = \mathcal{N}^{wBL} - \mathcal{N}^{wFL} = \sum_{m=1}^M \sum_{t=1}^T w_t (I_{tm}^{BL} - I_{tm}^{FL}).$$

Lastly, in order to study the effect of a positive trend in both tails we will use

$$D = \mathcal{N}^{wFU} - \mathcal{N}^{wFL} - \mathcal{N}^{wBU} + \mathcal{N}^{wBL} = \sum_{m=1}^M \sum_{t=1}^T w_t (I_{tm}^{FU} - I_{tm}^{FL} - I_{tm}^{BU} + I_{tm}^{BL}).$$

At the end of the chapter, a modification of the previous tests for M dependent series is proposed.

In Chapter 3, a statistical analysis of data is performed using the tools of Chapter 2 and graphical tools in order to find evidence of climate change. Such tools are implemented in the R *RecordTest* library. The series of daily maximum and minimum daily temperatures in the locations of Madrid, Daroca, Castellón, Tortosa, Izaña, Huelva and Cáceres are studied, focusing on Madrid as an example. First, the data transformation processes to obtain M independent series without seasonal component of length T are explained. After the data transformation, an analysis is carried out, which consists of studying the number of records of the series of maximum and minimum temperatures in Madrid. Also, at the end of the chapter, a summary for the rest of the locations is presented. The results obtained, both graphically and numerically, for all locations except for Cáceres, suggest the existence of a positive trend in the tails of the series, which encourages the presence of climate change.

Índice general

Resumen	III
Abstract	V
1. Teoría de Récorde	1
1.1. Nociones básicas	1
1.2. Variables indicadoras de récord I_t	2
1.3. Número de récorde N_t	2
2. Contrastes estadísticos	5
2.1. Contraste basado en un estadístico básico	5
2.2. Contraste basado en un estadístico con pesos	6
2.3. Elección de pesos	7
2.4. Contrastes basados en cuatro tipos de récorde	8
2.5. Contraste para M series dependientes	9
3. Análisis de datos	11
3.1. Datos	11
3.2. Análisis de Madrid	12
3.2.1. Análisis exploratorio de series de temperaturas diarias	12
3.2.2. Preparación de datos	13
3.2.3. Inferencia de series	14
3.3. Análisis de todas las series	18
3.3.1. Análisis temperatura máxima diaria en todo el año	18
3.3.2. Análisis temperatura mínima diaria en todo el año	20
3.3.3. Análisis por estaciones	20
3.4. Conclusiones	21
A. Análisis descriptivo previo	23
A.1. Datos incorrectos	23
A.2. Datos perdidos	23
A.2.1. Total NA's	23
A.2.2. NA's por década	24
B. Figuras y tablas	25
B.1. Castellón	25
B.2. Daroca	28
B.3. Huelva	31
B.4. Izaña	34
B.5. Tortosa	37
B.6. Cáceres	40
B.7. Madrid (resto)	43

B.8. Mapas	44
B.9. Tablas por estación	46
C. Código R	47
Bibliografía	61

Capítulo 1

Teoría de Récorde

Un récord es una medida que representa el valor más alto o bajo registrado dentro de una categoría. Existen numerosos campos en los que resulta interesante el estudio de récords. Por ejemplo, entre los campos más conocidos se encuentran la climatología y el mundo de los deportes, donde un estudio de los récords puede ayudar a entender los efectos de cambio climático o reflejar los límites del rendimiento del ser humano. A parte de estos, existen numerosos campos donde un estudio de récords proporciona información interesante y valiosa, como la hidrología, economía y medicina. En todas estas áreas, el estudio de récords permite comparar, analizar y comprender mejor el mundo que nos rodea.

Chandler (1952)[1] introdujo el estudio de los valores de récords y documentó muchas de sus propiedades básicas. En este capítulo se abordarán los conceptos básicos del *Modelo Clásico de Récorde* y nos centraremos en las variables de mayor relevancia para estudiar la ocurrencia de récords. En particular nos centraremos en las variables indicadoras de récords y el número de récords hasta un instante. Específicamente, se mostrará como estas variables muestran una distribución que no depende de la distribución de la serie original. Esta propiedad resulta de gran importancia ya que permitirá crear herramientas no paramétricas para el análisis de datos.

1.1. Nociones básicas

A lo largo de todo el capítulo, se considerará una secuencia de v.a.i.i.d. $(X_t)_{t \geq 1}$ con función de distribución F . Además, la distribución se supondrá continua para evitar la posibilidad de empates.

Una observación X_j será llamada *récord superior* si su valor excede el de sus precedentes, es decir,

$$X_j > \max_{i < j} (X_i).$$

En virtud de esta definición, se considera que la primera observación X_1 siempre es un récord trivial.

A partir de esta definición se pueden definir otras variables relacionadas con los récords. Por un lado, se define la sucesión de v.a. *indicadoras de récord* $(I_t)_{t \geq 1}$ como

$$I_t = \begin{cases} 1 & \text{si } X_t \text{ es récord} \\ 0 & \text{si } X_t \text{ no es récord.} \end{cases}$$

A su vez, se introduce el *número de récords* $(N_t)_{t \geq 1}$, donde N_t indica el número de récords que se producen en la secuencia X_1, \dots, X_t . Nótese que esta variable se puede definir a partir de las variables indicadoras de récords como

$$N_t = \sum_{j=1}^t I_j. \quad (1.1)$$

También, se encuentran los *instantes de ocurrencia de récords* $(L_i)_{i \geq 0}$. Estas indican el instante t en que se produce un récord y se definen como $L_0 = 1$ con probabilidad 1 y para $i \geq 1$, $L_i = \min\{t : X_t >$

$X_{L_{i-1}}\}$. A partir de estas, se define la secuencia de *valores de récord* $(R_i)_{i \geq 0}$ como $R_i = X_{L_i}$, donde R_0 se refiere al récord trivial X_1 . Todas las variables anteriores forman parte de lo que se conoce como el *Modelo Clásico de Récorde*, el cual se define a continuación.

Definición 1. Modelo Clásico de Récorde (MCR). Marco teórico que estudia la aparición de récorde en una secuencia de variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.)

Se puede señalar que bajo la aplicación de una función creciente g sobre la serie (X_t) , sus respectivas series de variables indicadoras de récord, número de récorde e instantes de ocurrencia de récorde se mantienen iguales a las de la serie original (X_t) .

En las siguientes secciones se demostrará la independencia de distribución de las variables I_t y N_t de la distribución de las variables X_t , la cual es una propiedad que resulta de gran importancia pues permitirá crear herramientas no paramétricas para el análisis de ocurrencia de récorde.

Además de récorde superiores, también existen los *récorde inferiores*. Para la serie $(X_t)_{t \geq 1}$, X_j es un récord inferior si $X_j < \min_{i < j} (X_i)$. Todas definiciones anteriores y resultados próximos se pueden aplicar y obtener de forma análoga a la de los récorde superiores debido a que $\min_{i < j} (X_i) = -\max_{i < j} (-X_i)$.

En las siguientes secciones se estudiarán las variables I_t y N_t para récord superiores.

1.2. Variables indicadoras de récord I_t

En esta sección estudiaremos algunas de las propiedades de las variables indicadoras de récorde superiores. Todas las variables de (I_t) son independientes entre sí y, además, cada una de ellas presenta una distribución de Bernoulli con probabilidad de éxito $1/t$. Esto queda reflejado y demostrado en la siguiente proposición.

Proposición 1.1. Sean X_1, X_2, \dots v.a.i.i.d. con función de distribución continua F . Entonces las variables indicadoras de récord I_1, I_2, \dots son mutuamente independientes y

$$P(I_t = 1) = \frac{1}{t}, \quad t = 1, 2, \dots$$

Demostración. (Nevzorov (2001, pp 58))[5] Sea $n \geq 2$ y $M_n = \max(X_1, X_2, \dots, X_n)$. Como X_1, X_2, \dots son v.a.i. con la misma distribución, se tiene que $P(X_i = M_n) = P(X_j = M_n)$, $i \neq j, i, j = 1, 2, \dots, n$. Es decir, para v.a.i.i.d. la probabilidad de que el máximo sea cualquiera de ellas es la misma. Entonces,

$$P(I_n = 1) = P(X_n = M_n) = P(X_k = M_n), \quad k = 1, 2, \dots, n.$$

Por continuidad de la función de distribución se tiene que para todo $i \neq j$, $P(X_j \neq X_i) = 0$. Entonces,

$$1 = P(M_n = X_1) + \dots + P(X_n = M_n) = nP(X_n = M_n).$$

Por tanto, se tiene que

$$P(I_n = 1) = \frac{1}{n}.$$

□

1.3. Número de récorde N_t

En esta sección se tratarán propiedades de las variables N_t . En especial se verá como cumplen el *Teorema Central del Límite (TCL)* de Lyapunov.

Haciendo uso de la expresión (1.1) y la Proposición 1.1, podemos calcular la esperanza y varianza de las variables N_t .

$$\mu_t = E(N_t) = \sum_{j=1}^t \frac{1}{j}$$

$$\sigma_t^2 = \text{Var}(N_t) = \sum_{j=1}^t \left(\frac{1}{j} - \frac{1}{j^2} \right)$$

Para la demostración del siguiente teorema se hará uso de dos conceptos, los *números de Stirling de primera especie sin signo* y la *función generadora de probabilidad*.

Definición 2. Sea $z \in \mathbb{C}, n, k \in \mathbb{N}_0$. Los números de Stirling de primera especie sin signo son los coeficientes $|s(n, k)|$ que cumplen

$$(z + n - 1)_n = \sum_{k=0}^n |s(n, k)| z^k,$$

donde $(x)_n = x(x-1)\dots(x-n+1), n \in \mathbb{N}$ y $(x)_0 = 1$.

Definición 3. Sea X una v.a. discreta no negativa. La función generadora de probabilidad (fgp) de la v.a. X se define como $E(z^X), z \in \mathbb{C}$.

Proposición 1.2. Sea X una v.a. discreta no negativa. Sea $E(z^X)$ su función generadora de probabilidad. Entonces,

$$P(X = j) = \left[\frac{1}{j!} \frac{d^j E(z^X)}{dz^j} \right]_{z=0}, \quad j = 0, 1, \dots$$

Demostración. Para la demostración del resultado, basta aplicar la definición de esperanza y derivar. Sea $j \in \mathbb{N}$ y X una v.a. discreta no negativa. Sea $z \in \mathbb{C}$, la fgp de X viene dada por $E(z^X) = \sum_{i \in \mathbb{N}} z^i P(X = i)$. Derivando j veces sobre z , se tiene que

$$\frac{d^j E(z^X)}{dz^j} = j! P(X = j) + (j+1)j \dots 2z P(X = j+1) + \dots$$

Tomando $z = 0$, se tiene el resultado. □

Una vez, definidas las herramientas y nociones anteriores, se puede demostrar el teorema que da la distribución exacta de las variables N_t .

Teorema 1.3. Sea $(X_t)_{t \geq 1}$ una serie de v.a.i.i.d. con función de distribución F continua. Sea $(N_t)_{t \geq 1}$ su serie de número de récords. Entonces la distribución de la variable N_t es,

$$P(N_t = i) = \frac{|s(t, i)|}{t!}, \quad t = 1, 2, \dots, \quad i = 1, 2, \dots, t.$$

Demostración. Para calcular la distribución de N_t se hará uso de su función generadora de probabilidad y los números de stirling de primera especie sin signo. Por la expresión (1.1) y la Proposición 1.1 se puede escribir la fgp de la v.a. N_t como la siguiente expresión.

$$E(z^{N_t}) = \prod_{j=1}^t E(z^{I_j}) = \prod_{j=1}^t \left(\frac{1}{j} z + 1 \left(1 - \frac{1}{j} \right) \right) = \frac{1}{t!} \prod_{j=1}^t (z + j - 1) = \frac{1}{t!} (z + t - 1)_t.$$

Por otro lado, se tiene,

$$E(z^{N_t}) = \sum_{j=1}^t z^j P(N_t = j).$$

Entonces, igualando ambas expresiones se tiene que,

$$P(N_t = i) = \text{coeficiente de } z^i \text{ en } \frac{1}{t!} (z + t - 1)_t = \frac{|s(t, i)|}{t!}.$$

□

Aunque el resultado anterior da las probabilidades exactas, su cálculo resulta complicado. En la práctica resulta más sencillo utilizar distribuciones aproximadas más fáciles de calcular. En particular, se verá como estas variables pueden ser aproximadas a una v.a. normal estándar. Para obtener dicha aproximación, se hará uso de la condición de Lyapunov.

Proposición 1.4. Condición de Lyapunov. Sea $X = \sum_{i=1}^n X_i$, con X_i v.a.i. Sea $\mu_i = E(X_i)$, $\sigma^2 = \text{Var}(X_i)$ y $s_n^2 = \sum_{i=1}^n \sigma_i^2$. Si existe $\delta > 0$ tal que

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E|X_i - \mu_i|^{2+\delta} = 0,$$

entonces,

$$\frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i) \rightarrow N(0, 1).$$

Demostración. Billingsley (1995, pp 359-362) [17]. □

Teorema 1.5. Sea $(X_t)_{t \geq 1}$ una serie de v.a.i.i.d con función de distribución F continua. Sea $(N_t)_{t \geq 1}$ su serie de número de récords. Entonces,

$$\frac{N_t - \mu_t}{\sigma_t} \xrightarrow{D} N(0, 1).$$

Donde $\mu_t = E(N_t) = \sum_{j=1}^t \frac{1}{j}$ y $\sigma_t^2 = \text{Var}(N_t) = \sum_{j=1}^t \left(\frac{1}{j} - \frac{1}{j^2} \right)$.

Demostración. Por (1.1) y la Proposición 1.1, N_t se puede expresar como una suma de v.a. Bernoulli independientes I_i con probabilidad de éxito $p_i = 1/i$, $i \in \mathbb{N}$. Nótese que $E(I_i) = \frac{1}{i} < \infty$ y $\text{Var}(I_i) = \frac{1}{i} - \frac{1}{i^2} \leq 1/4$, para todo $i \in \mathbb{N}$. Para lograr el resultado veremos que se cumple la Condición de Lyapunov.

Primero, notar que para el caso de v.a. Bernoulli, se tiene que

$$E|I_i - p_i|^3 = p_i(1 - p_i)^3 + (1 - p_i)p_i^3 < p_i(1 - p_i) = \text{Var}(I_i).$$

Sea $s_n^2 = \sum_{i=1}^n \text{Var}(I_i)$. Entonces,

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^3} \sum_{i=1}^n E \left| I_i - \frac{1}{i} \right|^3 \leq \lim_{n \rightarrow \infty} \frac{1}{s_n} = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{\sum_{i=1}^n \left(\frac{1}{i} - \frac{1}{i^2} \right)}} = 0.$$

Donde la última desigualdad se sigue de que la suma parcial $S_n = \sum_{i=1}^n \left(\frac{1}{i} - \frac{1}{i^2} \right) \approx \log(n) + \gamma - \sum_{i=1}^n \frac{1}{i^2}$, donde $\gamma = 0,5772\dots$ es la constante de Euler-Marscheroni, tiende a infinito cuando $n \rightarrow \infty$.

Por lo tanto se cumplen las condiciones necesarias para el TCL de Lyapunov y se tiene que

$$\frac{\sum_{i=1}^t (I_i - \frac{1}{i})}{\sqrt{\sum_{i=1}^t \text{Var}(I_i)}} = \frac{N_t - \mu_t}{\sigma_t} \xrightarrow{D} N(0, 1).$$

□

Como se ha mencionado al final de la Sección 1.1, estos resultados también se cumplen para el caso de récords inferiores.

Capítulo 2

Contrastes estadísticos

El problema de interés de este capítulo consiste en presentar diferentes contrastes de hipótesis basados en el Modelo Clásico de Récords (Definición 1) que permitan detectar desviaciones de estacionariedad en las colas de una distribución, en concreto en la ocurrencia de récords.

Este problema puede manifestarse en numerosos contextos como las olas del mar, la velocidad del viento, los precios de la electricidad y el clima. En el ámbito climático, aunque existen numerosos estudios que evidencian un aumento en la media de la temperatura global, es de gran interés analizar los posibles efectos del cambio climático en las colas de la distribución de la temperatura, es decir, en la ocurrencia de extremos y en particular, en la ocurrencia de récords. Por otro lado, en el sector eólico, dicho análisis puede influir en la ubicación y diseño de parque eólicos con el fin de optimizar su eficiencia.

La mayor dificultad para este problema suele ser la escasez de datos presentes en las cola de las series. En este capítulo se mostrarán diferentes contrastes de hipótesis, aumentando la potencia de dichos contrastes progresivamente.

Los contrastes que se explican son no paramétricos y se pueden aplicar a series correladas con comportamiento estacional tras una transformación previa de los datos. En dicha transformación, se obtendrán M series independientes de longitud T . Al final de capítulo, se aborda el caso en el que se deba analizar M series dependientes.

2.1. Contraste basado en un estadístico básico

En este capítulo se asumirán $M \geq 1$ series continuas mutuamente independientes de longitud T , $(X_{t1}), (X_{t2}), \dots, (X_{tM})$. En el Capítulo 3 se explorará como obtenerlas a partir de una sola serie. A su vez, se definen las series de sus variables indicadoras de récord $(I_{t1}), (I_{t2}), \dots, (I_{tM})$ y número de récords $(N_{t1}), (N_{t2}), \dots, (N_{tM})$. También se definirá el número de récords en el instante t en las M series como $S_t = I_{t1} + I_{t2} + \dots + I_{tM}$.

Bajo el MCR, las series son series i.i.d. y se comportan de manera estacionaria. Entonces, para nuestro problema de interés, el contraste tendrá como hipótesis nula,

$$H_0 : p_{tm} = \frac{1}{t}, \quad t = 1, \dots, T \quad m = 1, \dots, M, \quad (2.1)$$

donde $p_{tm} = P(I_{tm} = 1)$. Para analizar si existen desviaciones respecto al comportamiento estacionario, la hipótesis alternativa será que dicha probabilidad de ocurrencia de récord sea distinta a la anterior.

$$H_1 : p_{tm} \neq \frac{1}{t}, \quad \text{para al menos un } t = 1, \dots, T, \quad m = 1, \dots, M \quad (2.2)$$

En particular, si la hipótesis alternativa es que haya efectos de una tendencia creciente en la serie de ocurrencia de récords, esta sería $p_{tm} > 1/t$. Para detectar una tendencia decreciente se utilizaría $p_{tm} < 1/t$.

Tras presentar las hipótesis y series que se van a utilizar, se propone un contraste basado en un estadístico básico con el fin de detectar evidencias en contra de un comportamiento estacionario. Dicho

estadístico proporciona el número total de récords que se producen en las M series de longitud T . Este se define de la siguiente forma.

$$\mathcal{N} = \sum_{m=1}^M N_{Tm} = \sum_{t=1}^T \sum_{m=1}^M I_{tm} = \sum_{t=1}^T S_t.$$

Donde $S_t = \sum_{m=1}^M I_{tm}$ es el número de récords en el instante t .

El Teorema 1.5 nos indica la distribución asintótica normal que siguen las variables N_t , cuando t tiende a infinito. Entonces, bajo la hipótesis nula, \mathcal{N} es la suma de M variables asintóticamente normales cuando $T \rightarrow \infty$. Luego, el estadístico \mathcal{N} es asintóticamente normal cuando $T \rightarrow \infty$. Por otro lado, las variables S_t , por ser la suma de M v.a.i.d., también son asintóticamente normales cuando $M \rightarrow \infty$ por el Teorema Central del Límite. Entonces, \mathcal{N} también es asintóticamente normal cuando $M \rightarrow \infty$. Esto implica que la distribución del estadístico es asintóticamente normal cuando T y/o M tienden a infinito. Además la esperanza y la varianza del estadístico bajo la hipótesis nula son $e_{\mathcal{N}} = E(\mathcal{N}) = M \sum_{t=1}^T \frac{1}{t}$ y $v_{\mathcal{N}} = \text{Var}(\mathcal{N}) = M \sum_{t=2}^T \frac{1}{t} \left(1 - \frac{1}{t}\right)$ respectivamente.

A partir del estadístico \mathcal{N} , se crea un contraste de hipótesis. Sea $Z \sim N(0, 1)$ y \mathcal{N}_0 el valor observado de \mathcal{N} , el p -valor viene dado por

$$P\left(Z > \frac{\mathcal{N}_0 - e_{\mathcal{N}} - 0,5}{\sqrt{v_{\mathcal{N}}}}\right), \quad (2.3)$$

donde se ha añadido una corrección de continuidad.

El problema que presenta dicho contraste es que las series de récords aportan poca información sobre la muestra con la que se trabaja y por tanto, la potencia de dicho contraste es baja. En las siguientes secciones se presentan contrastes con mayor potencia.

2.2. Contraste basado en un estadístico con pesos

En la sección anterior se mostró un primer contraste de hipótesis para detectar desviaciones de estacionariedad. En esta capítulo se propone una modificación del anterior para aumentar la potencia del contraste.

Por la Proposición 1.1, la probabilidad de que ocurra un récord en el instante t , bajo la hipótesis nula es $1/t$. Dicha probabilidad disminuye conforme t aumenta, es decir, para valores de t altos, la probabilidad de que ocurra un récord es menor. Basándose en esta noción sobre las variables indicadoras de récord I_t , Diersen y Trenkler (1996)[10] dedujeron un estadístico más potente el cual consistía asignar pesos w_t , crecientes en t , a las variables S_t . De esta manera, si se produce un récord para un valor alto de t , este proporcionará más evidencias de no estacionariedad. Dicho estadístico se define como,

$$\mathcal{N}^w = \sum_{t=1}^T w_t S_t$$

De forma similar al estadístico \mathcal{N} , bajo la hipótesis nula \mathcal{N}^w es asintóticamente normal cuando $M \rightarrow \infty$. Por el contrario, cuando $T \rightarrow \infty$ no existe una aproximación normal debido a que la suma $\sum_{t=1}^T w_t I_{tm}$ no es asintóticamente normal (Diersen y Trenkler 2001).

La media y varianza de este estadístico bajo la hipótesis nula son

$$e_{\mathcal{N}^w} = E(\mathcal{N}^w) = M \sum_{t=1}^T w_t \frac{1}{t}.$$

$$v_{\mathcal{N}^w} = \text{Var}(\mathcal{N}^w) = M \sum_{t=2}^T w_t^2 \frac{1}{t} \left(1 - \frac{1}{t}\right).$$

Con esta información, se define un contraste de hipótesis más potente al anterior (2.3). Sea $Z \sim N(0, 1)$ y \mathcal{N}_0^w el valor observado de \mathcal{N}^w . El p -valor obtenido es

$$P\left(Z > \frac{\mathcal{N}_0^w - e_{\mathcal{N}^w} - 0,5}{\sqrt{v_{\mathcal{N}^w}}}\right).$$

2.3. Elección de pesos

En la sección anterior se presenta un estadístico con pesos w_t proporcionales a t . Diersen y Trenkler recomendaron pesos $w_t = t - 1$. En esta sección se verá como los pesos que proporcionan el contraste insesgado localmente más potente son $w_1 = 0, w_t = t^2/(t - 1), t = 2, \dots, T$.

Dadas M series independientes de longitud T de las variables (I_{tm}) , se propone un contraste basado en la función score. Recordar que la función score es el gradiente del logaritmo de la función de verosimilitud respecto a los parámetros de la distribución. La función de verosimilitud de las series es

$$\mathcal{L}(\mathbf{p}_{\mathbf{T}\mathbf{M}}|\mathbf{I}_{\mathbf{T}\mathbf{M}}) = \prod_{m=1}^M \prod_{t=2}^T p_{tm}^{I_{tm}} (1 - p_{tm})^{1-I_{tm}}, \quad (2.4)$$

donde $\mathbf{p}_{\mathbf{T}\mathbf{M}} = (p_{T1}, \dots, p_{TM})$ con $p_{Tm} = (p_{1m}, \dots, p_{Tm})$. Análogamente, $\mathbf{I}_{\mathbf{T}\mathbf{M}} = (I_{T1}, \dots, I_{TM})$ con $I_{Tm} = (I_{1m}, \dots, I_{Tm})$. Nótese que el factor para $t = 1$ es igual a 1

Si las series no se suponen i.i.d., las probabilidades de récord en el tiempo t en las M series pueden ser distintas. Debido a esto, hay un total de $M(T - 1)$ probabilidades p_{tm} distintas. Se observa que el número total de parámetros desconocidos es entonces igual al número de observaciones, y además, estos no pueden ser estimados.

King y Wu (1997)[13] proponen un método general para construir el contraste insesgado más potente con respecto a la media para contrastes con hipótesis alternativas unilaterales. Dicho contraste presenta propiedades óptimas de potencia cuando se trabajan con parámetros no molestos, el cual es nuestro caso. Un parámetro molesto es aquel parámetro no especificado que debe tenerse en cuenta en los contrastes de hipótesis. El estadístico del contraste resultante es la suma de los vectores score bajo la hipótesis nula.

En nuestro caso, bajo la hipótesis nula se tiene que $p_{tm} = 1/t, t = 1, \dots, T, m = 1, \dots, M$. Usando la función de verosimilitud (2.4) y esta hipótesis, el vector score, es el vector de longitud $M(T - 1)$ $\mathbf{q}_0 = (q'_{1,0}, \dots, q'_{M,0})$, donde

$$q_{m,0} = \left[\frac{t}{t-1} (tI_{tm} - 1) \right]_{t=2, \dots, T}.$$

La matriz información bajo la hipótesis nula, \mathcal{J}_0 es una matriz diagonal $M(T - 1) \times M(T - 1)$ con vector diagonal $\mathcal{J}'_d = (\mathcal{J}'_{d1}, \dots, \mathcal{J}'_{dM})$ donde cada subvector es

$$\mathcal{J}_{dm} = \left[\frac{t^2}{t-1} \right]_{t=2, \dots, T}.$$

Sea $\mathbf{1}_K$ el vector unidad de longitud K . Se considera el siguiente estadístico.

$$\mathcal{S} = \frac{\sum_{m=1}^M q_{m,0} \mathbf{1}_{T-1}}{\sqrt{\mathbf{I}_{M(T-1)} \mathcal{J}_0 \mathbf{1}_{M(T-1)}}} = \frac{\sum_{t=2}^T t(S_t - M)/(t-1)}{\sqrt{M \sum_{t=2}^T t^2/(t-1)}}. \quad (2.5)$$

Nótese que \mathcal{S} es una función lineal de las variables S_t , con pesos proporcionales a $t^2/(t - 1), t = 2, \dots, T$. Es decir, estamos tratando con un caso particular del estadístico \mathcal{N}^w definido en la Sección 2.2. Nuestro estadístico \mathcal{S} puede ser expresado como $\mathcal{S} = \sum_{t=2}^T w_t S_t + K$, con

$$w_t = \frac{t^2}{(t-1) \sqrt{M \sum_{t=2}^T t^2/(t-1)}} \text{ y } K = \frac{-\sum_{t=2}^T tM/(t-1)}{\sqrt{M \sum_{t=2}^T t^2/(t-1)}}.$$

La esperanza y varianza bajo la hipótesis nula de dicho estadístico son,

$$E(\mathcal{S}) = M \sum_{t=2}^T w_t \frac{1}{t} + K = 0.$$

$$Var(\mathcal{S}) = M \sum_{t=2}^T w_t^2 \left(\frac{1}{t} - \frac{1}{t^2} \right) = M \left(\frac{\sum_{t=2}^T (t^3 - t^2)/(t-1)^2}{M \sum_{t=2}^T t^2/(t-1)} \right) = 1.$$

Por ser una suma de funciones score, el estadístico \mathcal{S} es asintóticamente normal cuando $M \rightarrow \infty$. Debido a que su esperanza y varianza bajo la hipótesis nula son 0 y 1 respectivamente, el p -valor para la hipótesis alternativa es $P(Z > \mathcal{S}_0)$, donde \mathcal{S}_0 es el valor observado de \mathcal{S} y $Z \sim N(0, 1)$.

En conclusión, el contraste insesgado más potente resulta equivalente a utilizar pesos $w_t = t^2/(t-1)$ en el estadístico \mathcal{N}^w de la sección anterior.

2.4. Contrastes basados en cuatro tipos de récords

Un procedimiento que permite aumentar la potencia de los contrastes anteriores es obtener más información de la muestra uniendo cuatro tipos de récords. Esta idea fue sugerida por Foster y Stuart (1954)[12] y Diersen y Trenkler (1996)[10].

Los cuatro distintos tipos de récords son las series de récords superiores e inferiores hacia delante y hacia atrás. La serie hacia atrás es aquella que se obtiene cuando se revierten el orden de los términos de la serie para empezar a observar desde la última observación. Es decir, la serie X_T, X_{T-1}, \dots, X_1 .

El efecto de una tendencia creciente en las series hacia delante implica que la probabilidad de ocurrencia de récord superior e inferior es mayor y menor que $1/t$ respectivamente. En el caso de las series hacia atrás esta tendencia creciente se vuelve decreciente, haciendo que la probabilidad de récords superiores sea menor que $1/t$ y la probabilidad de récords inferiores mayor que $1/t$ (Véase Figura 2.1).

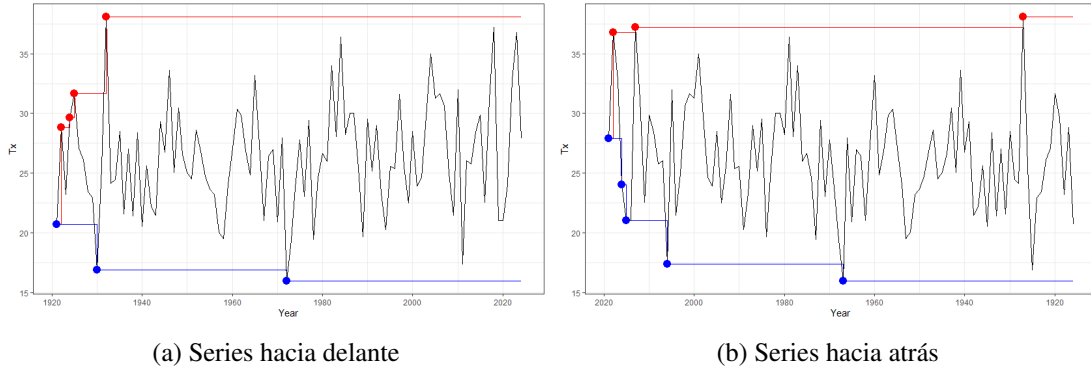


Figura 2.1: Series de temperaturas máximas del 11 de junio y récords superiores (rojo) e inferiores (azul) en la ciudad de Madrid entre 1920 y 2023.

Este procedimiento de unir diferentes series, nos permitirá crear estadísticos para contrastar diferentes hipótesis.

La primera hipótesis consiste en analizar el efecto de una tendencia creciente en la cola superior. Para ello, se considerarán las series indicadoras de récords superiores hacia delante I_t^{FU} y hacia atrás I_t^{BU} . Dichas series, bajo la hipótesis nula, son series i.i.d. y siguen una distribución *Bernoulli*($1/t$) (Proposición 1.1). El objetivo es crear un estadístico que aumente conforme t aumenta para detectar evidencias en contra de un comportamiento estacionario. Como se ha explicado previamente, bajo el efecto de una tendencia creciente, la probabilidad de éxito de las series hacia delante es mayor que $1/t$ y la de las series hacia atrás menor que $1/t$. Por esta razón, para obtener dicho estadístico se realizarán combinaciones lineales con I_t^{FU} y $-I_t^{BU}$. El estadístico resultante para analizar la no estacionariedad en la cola superior se define como,

$$U = \mathcal{N}^{wFU} - \mathcal{N}^{wBU} = \sum_{m=1}^M \sum_{t=1}^T w_t (I_{tm}^{FU} - I_{tm}^{BU}). \quad (2.6)$$

De manera análoga, utilizando las series indicadoras de récords inferiores hacia delante I_t^{FL} y hacia atrás I_t^{BL} , se obtiene un estadístico para contrastar la hipótesis de no estacionariedad en la cola inferior.

En este caso, la variable con signo negativo sera la serie hacia delante I_t^{FL} .

$$L = \mathcal{N}^{wBL} - \mathcal{N}^{wFL} = \sum_{m=1}^M \sum_{t=1}^T w_t (I_{tm}^{BL} - I_{tm}^{FL}). \quad (2.7)$$

Otra hipótesis que se puede contrastar con este procedimiento, es la no estacionariedad en ambas colas. En este caso se realizan combinaciones lineales con todas las series indicadoras de récords disponibles, $I_t^{FU}, I_t^{BU}, I_t^{FL}, I_t^{BL}$. De forma similar a los anteriores se deduce el siguiente estadístico.

$$D = \mathcal{N}^{wFU} - \mathcal{N}^{wFL} - \mathcal{N}^{wBU} + \mathcal{N}^{wBL} = \sum_{m=1}^M \sum_{t=1}^T w_t (I_{tm}^{FU} - I_{tm}^{FL} - I_{tm}^{BU} + I_{tm}^{BL}). \quad (2.8)$$

Se puede probar que los estadísticos U, L y D , bajo la hipótesis nula, son asintóticamente normales cuando $M \rightarrow \infty$. La esperanza de dichos estadísticos bajo la hipótesis nula es 0. Para la obtención de los p -valores correspondientes a cada contraste se procede de la forma usual. En la aplicación del Capítulo 3 se utilizaran estos tres contrastes para detectar desviaciones respecto a la estacionariedad en series de temperaturas máximas y mínimas diarias usando pesos $w_1 = 0, w_t = t^2/(t-1), t = 2, \dots, T$.

2.5. Contraste para M series dependientes

Los contrastes anteriores resultan de gran potencia cuando el número de series independientes M es mayor o igual que 15. La distribución asintótica de los estadísticos previos recae sobre la suposición de que las M series son independientes. Para evitar esta restricción, se propone un contraste de hipótesis alternativo en el que los p -valores de los estadísticos se calculan usando técnicas de permutación. En nuestro caso, en el Capítulo 3, se verá como utilizaremos este método para estudiar el comportamiento de series de temperaturas en las distintas estaciones del año.

Los contraste de permutación consisten en calcular todos los posibles valores del estadístico del contraste haciendo reorganizaciones de los datos. Estos contrastes dependen de la suposición de intercambiabilidad bajo la hipótesis nula (Welch 1990). Una muestra es intercambiable si cualquier permutación de esta sigue teniendo la misma distribución conjunta.

En nuestro caso, en los contraste de récords, contamos con T observaciones de un vector de longitud M , $(X_{t1}, X_{t2}, \dots, X_{tM})$. Bajo la hipótesis nula, las T observaciones del vector son independientes con la misma distribución, entonces, las filas de nuestros datos son intercambiables.

Si las observaciones son intercambiables bajo la hipótesis nula y consideramos todas las combinaciones posibles para calcular el p -valor, entonces, el contraste resultante tiene el mismo nivel de significancia. El p -valor es calculado como la proporción de muestras cuyo valor del estadístico es mayor o igual que el valor del estadístico observado. Se estima que un total de 10000 permutaciones da una buena aproximación del p -valor.

Capítulo 3

Análisis de datos

En este capítulo se realiza un análisis de datos de temperaturas máximas y mínimas diarias con el fin de analizar si existe efecto del calentamiento global en las colas de la distribución de series de temperaturas máximas y mínimas diarias de todas las series centenarias disponibles en la bases de datos European Climate Assessment & Dataset (ECA&D). Para el análisis estadístico se utilizan las herramientas de inferencia estadística basadas en récords explicadas en el Capítulo 2. Aunque estas herramientas se encuentran implementadas en la librería de R *RecordTest*[8], se han desarrollado funciones las cuales se pueden encontrar en el Apéndice C.

Recordar que para poder aplicar los contrastes basados en las distribuciones asintóticas, es necesario contar con M series independientes de longitud T . Las series con las que se va a trabajar muestran una fuerte componente estacional. Para deshacernos de dicha estacionalidad se explicarán los distintos métodos de transformación utilizados para la extracción de M series independientes de longitud T a partir de una única serie.

Posteriormente, se detallarán los pasos del análisis gráfico y numérico de cada serie, mostrando como ejemplo la serie de Madrid. Por último, se mostrará un resumen de los resultados de todas series analizadas.

3.1. Datos

En este capítulo se van a analizar series centenarias de la base de datos ECA&D de temperaturas máximas y mínimas diarias en España. Específicamente los datos provienen de observatorios localizados en Castellón, Madrid (Retiro), Daroca, Cáceres, Huelva, Izaña (Tenerife) y Tortosa. Los datos están medidos en grados Celsius (°C) con precisión de un decimal. La tabla 3.1 muestra los detalles de cada serie que se va a analizar. Dichos detalles coinciden para las series de ambos tipos de temperatura.

Estación	Año inicio-final	Nº obs	Nº NAs	Latitud	Longitud
Castellon	1920-2023	37960	397	39°57'00" N	00°04'17" W
Madrid	1920-2023	37960	175	40°24'43" N	03°40'41" W
Daroca	1920-2023	37960	186	41°06'53" N	01°24'39" W
Caceres	1920-2023	37960	101	39°29'00" N	06°22'00" W
Huelva	1920-2023	37960	158	37°15'35" N	06°56'57" W
Izaña	1920-2023	37960	413	28°18'32" N	16°29'58" W
Tortosa	1905-2023	43435	318	40°49'14" N	00°29'29" E

Cuadro 3.1: Información series disponible ECA&D

España se trata de un territorio rico en tipos de clima distintos. Los tipos de climas predominantes son el clima mediterráneo, de montaña, oceánico y subtropical. En particular, para la clasificación de *Köppen-Geiger* Cáceres, Castellón, Huelva y Tortosa tiene un clima mediterráneo caracterizado por veranos

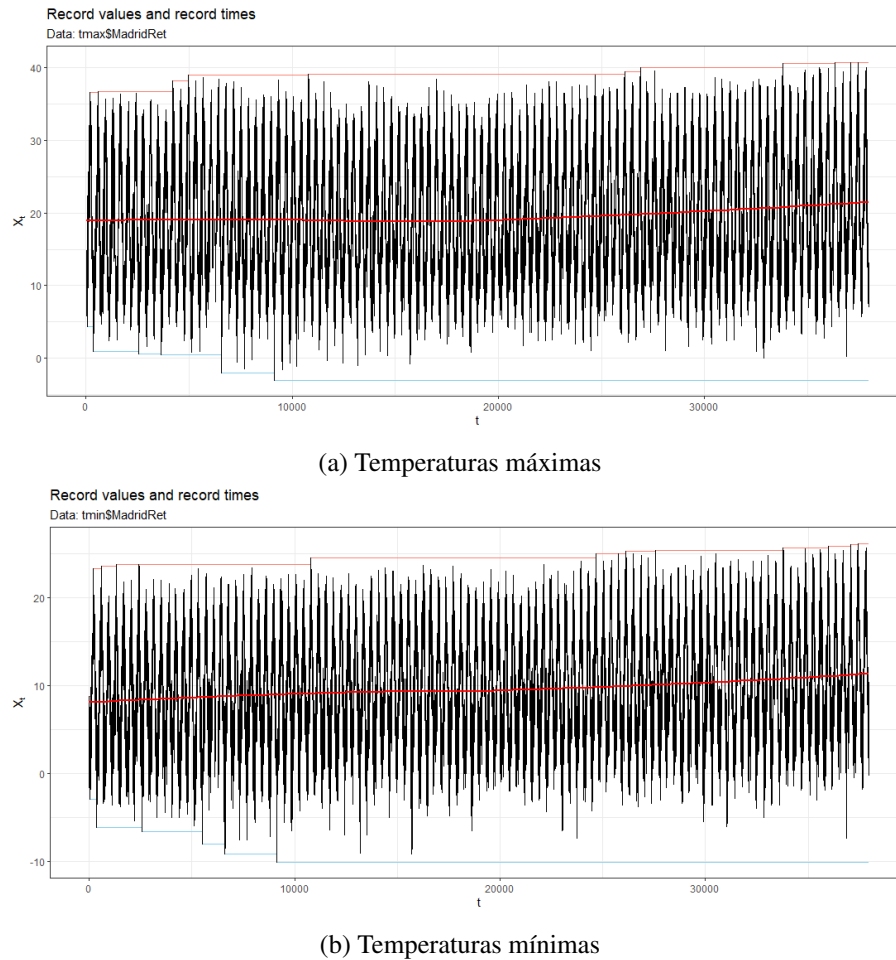


Figura 3.1: Temperaturas diarias en Madrid, España (1920-2023). LOESS (rojo sólido). Récorde superiores (rojo) e inferiores (azul).

cálidos y secos e inviernos lluviosos y fríos. Daroca presenta un clima semiárido frío caracterizado por inviernos fríos y secos, y con temperaturas medias anuales no superiores a 18 °C. Por último Izaña presenta un clima mediterráneo oceánico con un verano suave y seco.

3.2. Análisis de Madrid

En esta sección se van a desarrollar todos los pasos del análisis de cada serie. Se tomarán como series ejemplo las series de temperaturas máximas y mínimas diarias de Madrid. El proceso completo consiste en un análisis exploratorio, una preparación de datos y una inferencia de series en todo el año y por estaciones.

3.2.1. Análisis exploratorio de series de temperaturas diarias

El análisis exploratorio consiste en un estudio de tendencia en las series completas, y un estudio sobre el patrón estacional y evolución temporal.

Tendencia. Para observar la tendencia que presentan las series, se ilustra la serie al completo junto a una regresión local LOESS la cual nos indica dicha tendencia. A su vez, mostrando la serie al completo se puede observar una clara componente estacional. La Figura 3.1 muestra indicios de una tendencia positiva en temperaturas máximas y mínimas.

Patrón estacional y evolución temporal. Para apoyar estos indicios de una tendencia positiva se ilustra el comportamiento de las temperaturas en 3 periodos de 20 años. Para ello, se muestra, para cada

día del año, la temperatura media en esos periodos, con el fin de estudiar la posible evolución del patrón estacionario. Los periodos en los que se va a trabajar son de 1920 a 1939, de 1960 a 1979 y de 2000 a 2019. La Figura 3.2 muestra como ambos tipos de temperaturas siguen un patrón similar para los distintos periodos, así como la fuerte componente estacional. Sin embargo, se puede notar que conforme se avanza en el tiempo, hay un aumento de temperatura media, tanto para temperaturas máximas y mínimas, siendo el periodo mas reciente (azul) el que sobresale sobre las demás, mostrando indicios de existencia de una tendencia creciente.

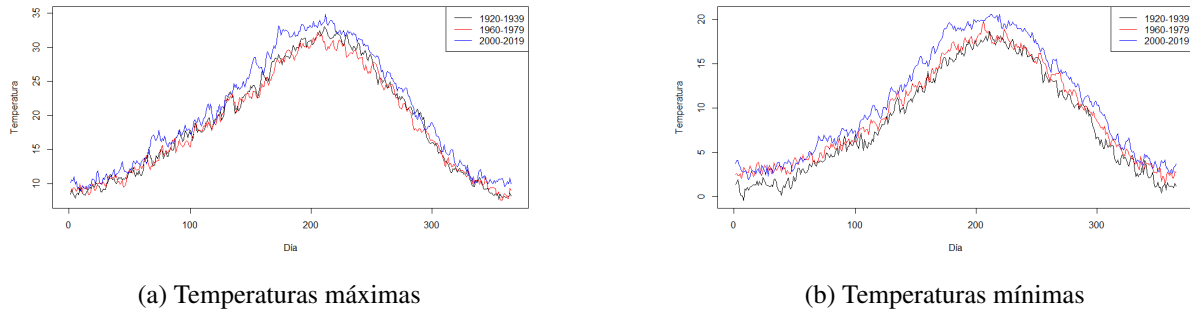


Figura 3.2: Temperaturas diarias medias en Madrid en diferentes periodos de 20 años, 1920 a 1939 (negro), 1960 a 1979 (rojo) y 2000 a 2019 (azul).

3.2.2. Preparación de datos

Tras el análisis exploratorio de los datos, se realiza una preparación de datos para la posterior aplicación de los contrastes explicados en el Capítulo 2. La preparación de datos consiste en considerar el comportamiento estacional, obtener M series independientes y romper los empates que puedan aparecer.

Series sin componente estacional. Las series de temperaturas diarias presentan una clara componente estacional. Un procedimiento habitual en análisis climáticos para tratar este problema consiste en separar las series por día del año. Dicha separación de datos implicará la creación de una matriz de 365 columnas y T filas, donde T es el número de años disponibles en la serie. De esta forma se obtienen 365 series temporales para cada localidad.

$$\begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,365} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,365} \\ \vdots & \vdots & & \vdots \\ X_{T,1} & X_{T,2} & \cdots & X_{T,365} \end{pmatrix}_{T \times 365}$$

Para cada serie temporal, el tiempo transcurrido entre observaciones es de un año, entonces se puede suponer que la correlación entre observaciones es nula. Debido al efecto estacional, la distribución de las 365 series será distinta, sin embargo, las T v.a. para cada serie pueden estar idénticamente distribuidas. Esto no supone un problema ya que los contraste que se van a aplicar no suponen que la distribución de M series temporales sea igual.

Series incorreladas entre sí. Tras la obtención de series sin componente estacional, es necesaria la obtención de M series incorreladas. El procedimiento usual para series temporales consiste en realizar el contraste de correlación de Pearson de manera iterativa. Primero, se selecciona la serie para el día k y se estudia la correlación con la serie $k + 1$. Si no existe correlación significativa se selecciona, y en caso contrario se estudia la correlación entre la serie k y $k + 2$. Este proceso se realiza hasta encontrar una serie $k + j$ incorrelada con k . En nuestro caso, empezamos con el primer día del año y buscamos un día con el que pueda ser incorrelado. Después consideramos ese nuevo día y volvemos a hacer lo mismo, y repetimos este proceso hasta llegar al último día.

Empates de temperaturas. El MCR asume que las series son v.a. continuas, entonces la probabilidad de que existan empates es cero. No obstante, en aplicaciones prácticas, esta suposición no se cumple debido a la precisión de los datos. Los datos con las que estamos trabajando tienen precisión de un decimal, entonces la probabilidad de encontrarnos con datos repetidos existe. Esto se suele deber al redondeo de datos.

Para romper dichos empates, sumaremos a cada dato un valor uniforme aleatorio entre $-0,05$ y $0,05$, de forma que las modificaciones son lo suficientemente pequeñas para no alterar la integridad de los datos.

Creación de series de récords. Tras realizar las transformaciones previas sobre la serie de temperaturas a estudiar, es necesario crear las series de récords incorreladas con la que se van a realizar los distintos contrastes. Recordar que dichas series son binarias, donde el valor 1 implica que se ha producido un récord. En el caso de Madrid, tras aplicar dicha transformación se obtienen 54 y 44 series independientes de longitud $T = 104$ para temperaturas máximas y mínimas respectivamente. En el caso de temperaturas máximas se encuentran 314 récords superiores y 266 inferiores. El valor esperado de récords en ambos casos es de 282. En el caso de temperaturas mínimas se observan 268 récords superiores, 217 inferiores y un valor esperado de 230. Como se puede observar en ambos tipos de temperaturas, los records superiores exceden al esperado, así como los inferiores están por debajo, mostrando posibles indicios de comportamiento no estacionario.

3.2.3. Inferencia de series

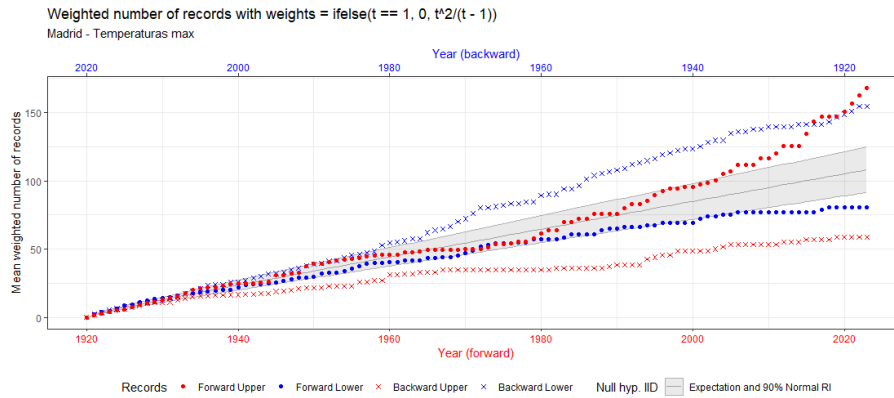
En esta sección, tras la previa preparación de datos, se detallan los pasos a realizar en el análisis de las series. Primero se verá un análisis de la distribución de las colas en todo el año, y posteriormente por estaciones.

Análisis en todo el año

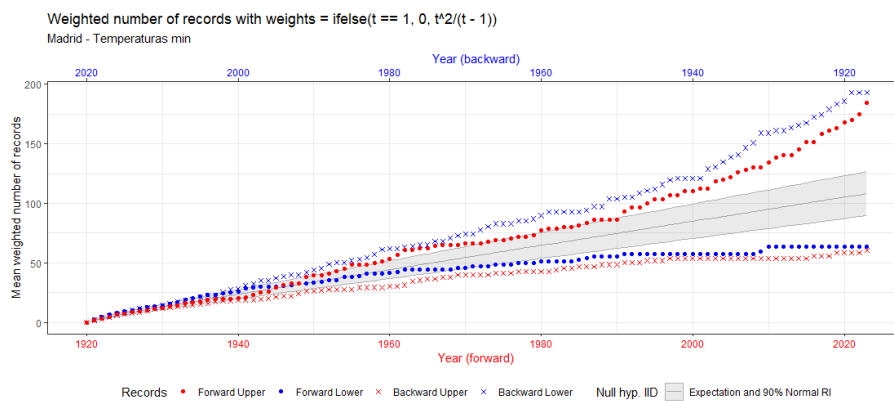
Primero, para detectar posibles desviaciones en estacionariedad, se analizará gráficamente la evolución del número de récords en el tiempo. Para ello, se ilustran en la misma gráfica el número de récords medio observados para cada instante en los cuatro tipos de récords. Los récords superiores serán mostrados en azul y los inferiores en rojo. Se usarán puntos para las series hacia delante y cruces para las series hacia atrás. Además, también se muestra el valor esperado bajo el MCR (series i.i.d.) junto a un intervalo de confianza del 90 % en gris con el fin de poder visualizar posibles desviaciones respecto a un comportamiento estacionario. Para una mayor claridad de resultados se utilizan pesos $w_0 = 0, w_t = t^2 / (t - 1), t = 2, \dots, T$.

La Figura 3.3a muestra como en las primeras décadas, el número de récords en temperaturas máximas sigue un comportamiento estacionario. Sin embargo, conforme avanzan los años se puede observar un desvío notable de las series hacia atrás, intensificándose al final de dicha serie. En el caso de las series hacia delante, esta desviación se produce a partir del año 2005 aproximadamente. Por el otro lado, en la Figura 3.3b se observa como, en comparación con las temperaturas máximas, el comportamiento no estacionario de las temperaturas mínimas es más significativo. Las series hacia atrás comienzan a desviarse muy pronto en el tiempo, y las series hacia delante muestran esta desviación de manera más notable a partir del año 1990.

A continuación se realizan contrastes de hipótesis para los estadísticos U (2.6), L (2.7) y D (2.8) con pesos $w_0 = 0, w_t = t^2 / (t - 1), t = 2, \dots, T$ para apoyar esta idea de comportamiento no estacionario en la distribución de las colas de las series. La hipótesis nula de dichos contrastes será que las series muestran comportamiento estacionario, es decir, son series i.i.d. en las cuales la probabilidad de que ocurra un récord en el instante t es igual a $1/t$. Para el caso de nuestros estadísticos esto se resume a que el valor esperado es igual a 0. Por otro lado, la hipótesis alternativa será que dichos estadísticos tienen valor mayor que 0, lo cual supondría un comportamiento no estacionario con tendencia creciente. Para estos contrastes se mostrarán los p -valores obtenidos. Se trabajará con los niveles de confianza habituales, 90 %, 95 % y 99 %.



(a) Temperaturas máximas



(b) Temperaturas mínimas

Figura 3.3: Media número de récords para temperaturas máximas y mínimas en Madrid desde 1920 a 2023.

Para temperaturas máximas diarias, el análisis de la cola superior (U) muestra un p -valor igual a $1,76e - 09$, mientras que para la cola inferior (L) el p -valor es igual a $3,19e - 05$. Entonces, para valores de confianza comunes, se rechaza la hipótesis nula, dando a entender que existe una tendencia positiva en ambos tipos de récord. Si se analizan ambas colas simultáneamente utilizando el estadístico D , se obtienen pruebas más significativas para rechazar la hipótesis nula. En este caso, el p -valor es igual a $2,04e - 12$.

En el caso de temperaturas mínimas diarias, el análisis de colas superiores e inferiores dan p -valores iguales a $1,15e - 09$ y $3,48e - 10$ respectivamente, mientras que el estudio de ambas colas da un p -valor igual a $9,38e - 18$. Por tanto, se rechaza la hipótesis nula de estacionariedad. Nótese, que para el caso de temperaturas mínimas, tanto gráficamente y numéricamente, esta tendencia positiva es más notable que para las temperaturas máximas.

Por último, para apoyar los resultados obtenidos en los contrastes de hipótesis anteriores, se muestra el valor medio observado del estadístico D a lo largo del tiempo junto a una banda gris de confianza del 90%. La Figura 3.4a exhibe como a partir del año 1980 existe un notable comportamiento no estacionario para temperaturas máximas diarias, mientras que para temperaturas mínimas, este comportamiento se hace notar a partir de 1960 aproximadamente (Figura 3.4b), llegando a la conclusión previamente explicada, los datos en temperaturas máximas y mínimas para Madrid en los últimos 104 años muestran un comportamiento claramente no estacionario en todo el año, siendo este mucho más apreciable para el caso de temperaturas mínimas, lo cual sugiere la existencia de cambio climático.

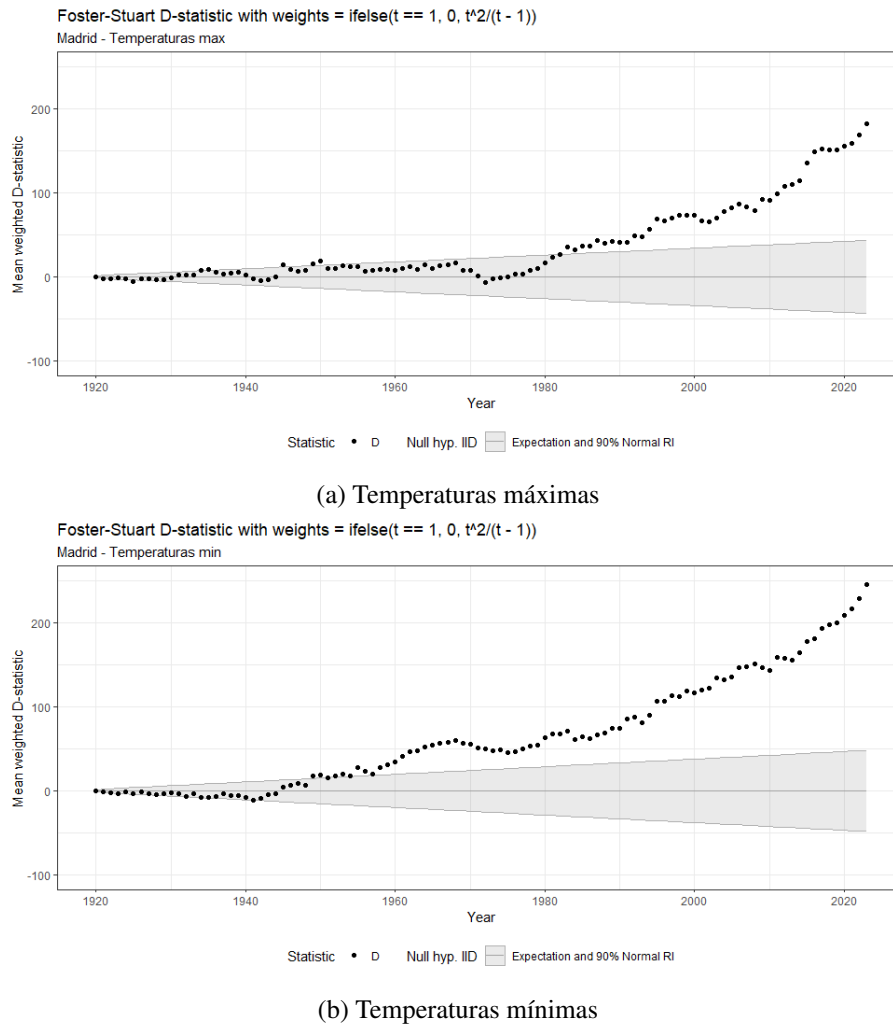


Figura 3.4: Valor medio estadístico D para temperaturas máximas y mínimas.

Análisis por estaciones del año

Otra hipótesis de interés a analizar es si los efectos del cambio climático son homogéneos a lo largo del año. Para ello se aplicarán los análisis anteriores separados para cada estación del año. En este caso, el número de series independientes M suele ser pequeño, lo que implica menor potencia de los contrastes. Entonces, es preferible aplicar la modificación explicada en la Sección 2.5 para M series dependientes. Dichos contrastes se realizarán con número de permutaciones igual a 5000.

Para cada estación, se calculan los p -valores de los estadísticos U , L y D como en la sección anterior. Además se muestran los gráficos correspondientes para el número de récords medio en los cuatro tipos de series para cada estación del año.

Estadísticos \ Estaciones	Invierno	Primavera	Verano	Otoño
Cola superior, U	0.024	0.006	4e-04	0.001
Cola inferior, L	0.238	0.374	0.008	4e-04
Ambas colas, D	0.045	0.034	2e-04	6e-04

Cuadro 3.2: p -valores de distintos estadísticos de temperaturas máximas diarias por estaciones

En la Tabla 3.2 se puede observar como para el caso de temperaturas máximas, las estaciones de verano y otoño muestran mayores evidencias en contra de la hipótesis nula. En los tres tipos de contras-

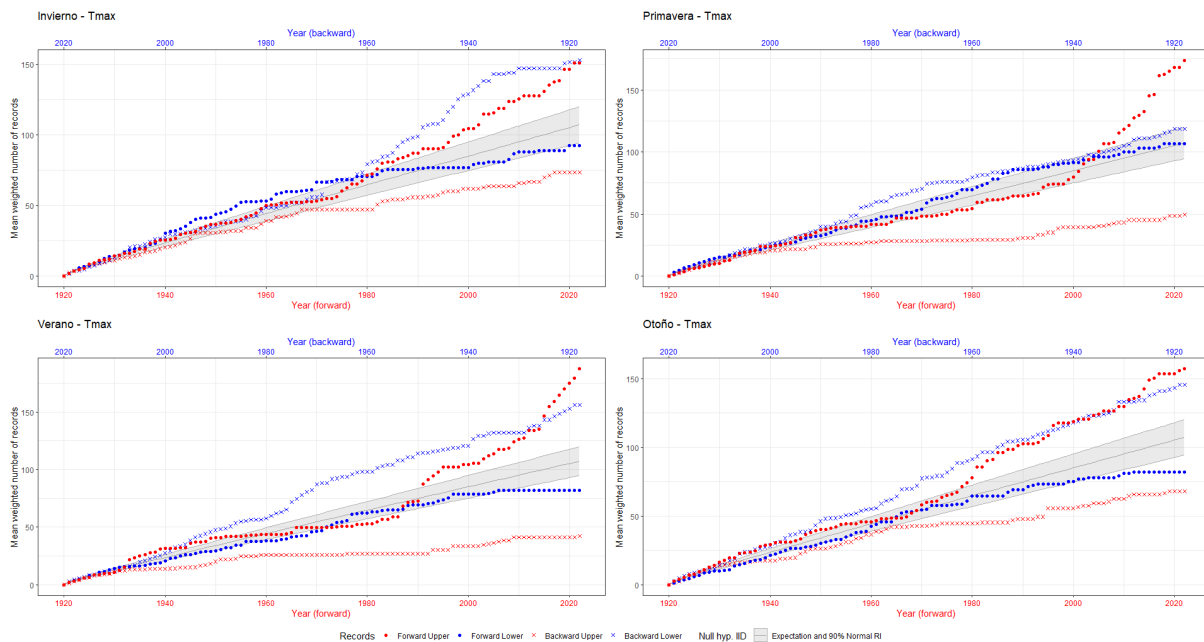


Figura 3.5: Media número de récords para temperaturas máximas por estación.

tes, la hipótesis nula es rechazada. Por el otro lado, para las estaciones de invierno y primavera, estas evidencias no son tan claras. Esto es debido a que la hipótesis nula para el contraste de colas inferiores no es rechazada. Si se observa la Figura 3.5, se contempla como para primavera, el comportamiento de las series correspondientes a récords inferiores (azul) es casi estacionario, lo cual provoca el p -valor obtenido. En invierno, este comportamiento estacionario se produce para la serie de récord inferiores hacia delante. Por el contrario, en verano y otoño se producen desvíos en la estacionariedad para todos los tipos de series, especialmente a partir de las últimas décadas, lo cual se ve reflejado en los p -valores obtenidos.

La Tabla 3.3 revela significativamente una tendencia positiva en todas estaciones del año para las temperaturas mínimas, especialmente en verano y otoño, donde los p -valores obtenidos son muy pequeños. En el caso de invierno y primavera, las evidencias de una tendencia positiva existen pero no de forma tan notable. La Figura 3.6 muestra como estos desvíos sobre la estacionariedad se producen muy pronto en el tiempo de manera creciente. En el caso de verano y otoño se puede observar como las primeras desviaciones se producen a partir de 1960 aproximadamente. En el caso de invierno y primavera, dichas desviaciones empiezan a ser notables a partir de 1980.

En conclusión, para temperaturas mínimas, se observa una tendencia positiva en temperaturas para todas estaciones del año, mientras que para temperaturas máximas, este comportamiento tan solo ocurre de forma significativa para las estaciones de verano y otoño. Comparando los p -valores obtenidos, así como las gráficas, las temperaturas mínimas en Madrid han sufrido un mayor aumento en relación a las temperaturas máximas. Independientemente de esta diferencia, para ambos tipos de temperaturas, se rechaza la hipótesis nula en el contraste del estadístico D , concluyendo en la existencia de efecto de un cambio climático.

Estadísticos \ Estaciones	Invierno	Primavera	Verano	Otoño
Cola superior, U	0.0366	0.0002	$< 2.2e-16$	$< 2.2e-16$
Cola inferior, L	0.0112	0.0008	$< 2.2e-16$	$< 2.2e-16$
Ambas colas, D	0.0064	$< 2.2e-16$	$< 2.2e-16$	$< 2.2e-16$

Cuadro 3.3: p -valores de distintos estadísticos de temperaturas mínimas diarias por estaciones

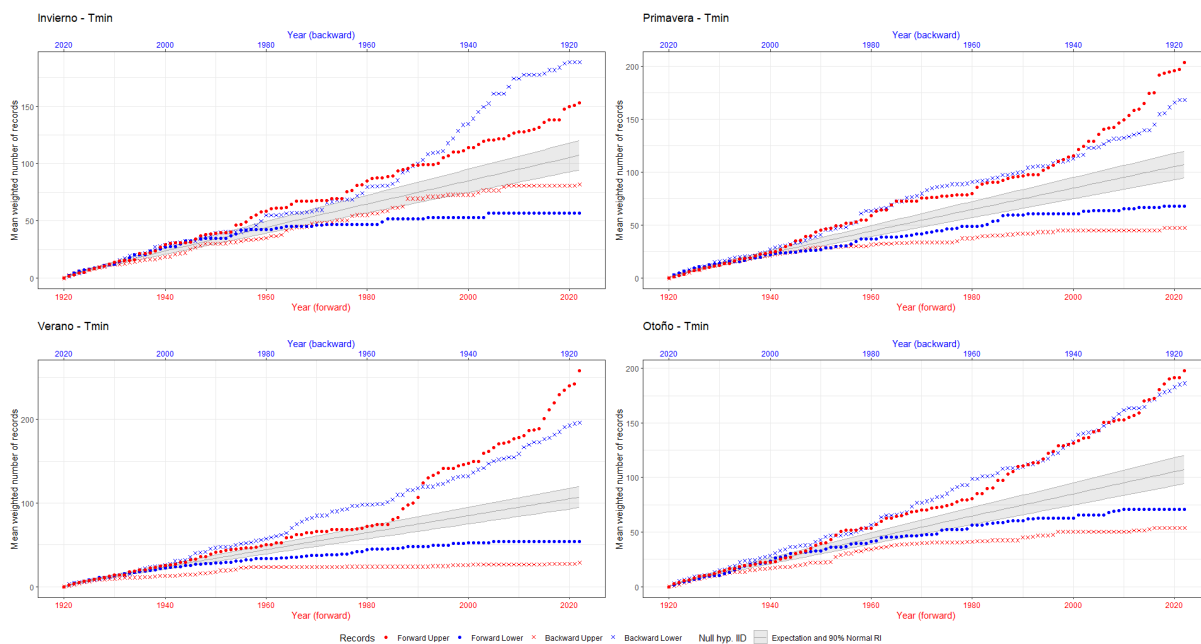


Figura 3.6: Media número de récords para temperaturas mínimas por estación.

3.3. Análisis de todas las series

En la sección anterior hemos realizado un estudio en profundidad para la ciudad de Madrid. En esta sección se tratarán de resumir los resultados más importantes para el resto de ciudades mencionadas, en particular se mostrarán los resultados para los contrastes basados en los estadísticos U , L y D . Recordar que la hipótesis nula de los siguientes contrastes es un comportamiento estacionario en las colas de la distribución de las series de ocurrencia de récords, mientras que la alternativa es la existencia de una tendencia creciente en estas.

3.3.1. Análisis temperatura máxima diaria en todo el año

Loc. Est.	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa	Madrid
Cola superior, U	0.005	5.95e-06	0.267	0.002	3.64e-07	3.62e-09	1.76e-09
Cola inferior, L	0.017	3.48e-07	0.006	9.00e-04	3.91e-04	0.09	3.19e-05
Ambas colas, D	5e-04	3.06e-11	0.014	1.22e-05	2.94e-09	2.93e-07	2.04e-12

Cuadro 3.4: P -valores temperaturas máximas por localidad.

La Tabla 3.4 resume los p -valores obtenidos para cada estadístico para temperaturas máximas en las localidades Castellón, Daroca, Cáceres, Huelva, Izaña y Tortosa. La primera hipótesis a estudiar es el efecto de una tendencia creciente en la cola superior. Se puede observar como los p -valores obtenidos para las ciudades de Daroca, Tortosa, Izaña y Madrid son de orden muy bajo, y por tanto, se rechaza la hipótesis de estacionariedad para cualquier nivel de significancia común. Por otro lado, en Castellón y Huelva, pese a haber evidencias de una tendencia creciente en la cola superior, esta no es tan notable como en las anteriores. Por último, para la ciudad de Cáceres la hipótesis nula no es rechazada debido al p -valor igual a 0,267. En cuanto al efecto en la cola inferior, se rechaza la hipótesis nula para la mayoría todas localidades, siendo los efectos de calentamiento global más apreciables en Daroca. Analizando de forma conjunta ambas colas, la existencia de desviaciones significativas en todas las series es clara, inclusive Cáceres, la cual presenta el p -valor más alto.

Adicionalmente, la Tabla 3.5 muestra el último año en que el se inician desviaciones crecientes respecto al comportamiento estacionario en las colas de la distribución. Aquellas celdas vacías indican que no existe una fecha clara de desviación con la escala utilizada para las gráficas individuales de los gráficos de U , L y D (Apéndice B). Dichas fechas ejercen de ayuda para comprender cuando el efecto de calentamiento global empezó a tener efecto sobre las colas de las distribuciones de las series de récords. Como se puede observar, dependiendo de la ubicación, los efectos han ido incrementando a un ritmo distinto. Se puede observar como para aquellas localidades con los mayores p -valores, el año de inicio es más reciente o incluso no es claro. Un comportamiento común entre localidades es que la aparición de desviaciones en ambas colas produce con anterioridad al resto, salvo en el caso de Tortosa y Huelva. Observando las fechas de manera conjunta, las desviaciones comienzan a aparecer a partir del año 1990.

Loc. Est.	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa	Madrid
Cola superior, U	2005	2010		2015	1965	1985	1995
Cola inferior, L	2020	1995	2020	1990	2010		1990
Ambas colas, D	2005	1990	2020	2005	1985	2005	1980

Cuadro 3.5: Fechas inicio desviaciones en temperaturas máximas por localidad.

Como se ha visto en los resultados anteriores, la serie de temperaturas máximas en Cáceres muestra los mayores p -valores. La Figura 3.7 muestra como, en general, el comportamiento de las series de récords de dicha serie se encuentran dentro de la zona de confianza, destacando las series de récords superiores, las cuales siguen un notable comportamiento estacionario. A su vez, destacar que la serie de récords inferiores hacia delante está por encima del valor esperado a lo largo de todo el tiempo, así como la serie hacia delante de récord superiores está por debajo a partir de 1960. También notar que el único tipo de récord que muestra desviaciones es el de récords inferiores hacia atrás, lo cual provoca el rechazo de la hipótesis nula para el contraste realizado para la cola inferior.

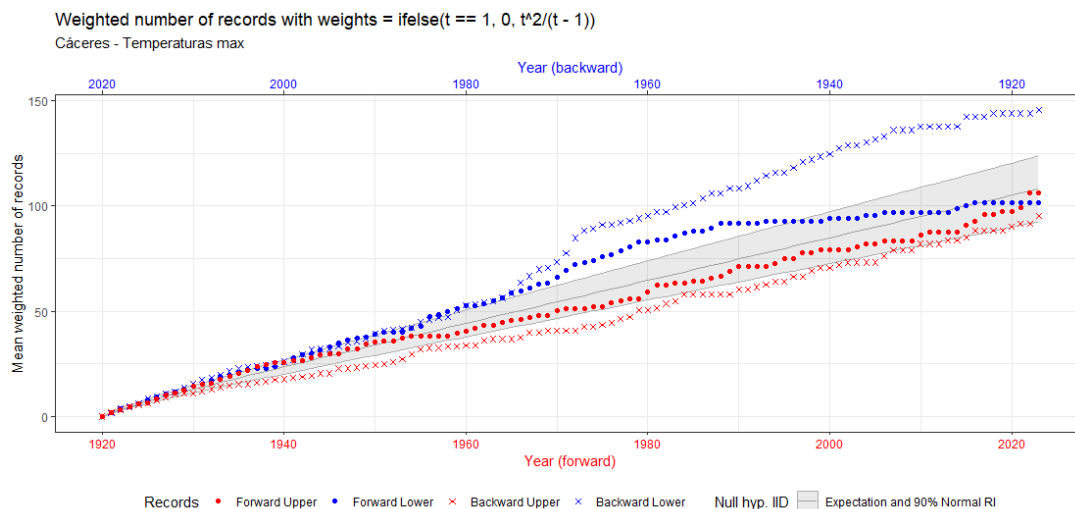


Figura 3.7: Media número de récords para temperaturas máximas.

Este comportamiento se puede intuir comparando el número de récords observados con el esperado. Tras la transformación de datos para obtener series sin componente estacional e incorreladas, obtenemos 61 series de longitud 104. El número de récords superiores, inferiores y esperado es 313, 335 y 319 respectivamente. Estos valores muestran las primeras sospechas de que la serie a estudiar no mostrará una tendencia creciente significativa en las colas. El número de récords inferiores es mayor que el de superiores, y además es mayor que el valor esperado. También, el número de récords superiores es inferior al esperado. Como se ha visto, en el caso de la cola superior, dichas sospechas son ciertas. Este

problema en la ciudad de Cáceres podría deberse a un error de medición.

3.3.2. Análisis temperatura mínima diaria en todo el año

Loc. Est.	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa	Madrid
Cola superior, U	0.003	0.044	0.094	0.046	6.19e-08	0.003	1.15e-09
Cola inferior, L	0.001	0.001	0.001	2e-04	2.71e-04	4.20e-14	3.48e-10
Ambas colas, D	2.35e-05	5.2e-04	0.001	1.3e-04	4.56e-10	3.41e-13	9.38e-18

Cuadro 3.6: P -valores temperaturas mínimas por localidad.

En la Tabla 3.6 se observa como, para temperaturas mínimas, todas las series muestran evidencias en contra de un comportamiento estacionario. Dichas desviaciones son más significativas cuando ambas colas y la cola inferior son consideradas. Por el otro lado, en la cola superior encontramos una mayor diferencia respecto a las otras dos, donde de nuevo, Cáceres presenta el mayor p -valor. Los p -valores de Daroca y Huelva, para la cola superior, se encuentran en la frontera, los cuales no serían significativos a un nivel $\alpha = 0,01$. Cabe destacar los p -valores obtenidos para las series de Tortosa y Madrid, las cuales presentan las evidencias más significativas de todo el estudio para la cola inferior y ambas colas.

La Tabla 3.7 indica los años de inicio de desviaciones para cada estudio. Apoyando lo mencionado anteriormente, se observa como la fecha de inicio de desviaciones en la cola superior para Cáceres, Huelva y Daroca no es clara. Por otro lado, destacar la serie de Tortosa cuando se consideran ambas colas, donde la tendencia se mantiene creciente desde 1915 de manera significativa. En el caso de Madrid, dicha tendencia aparece más tarde pero crece más rápido que en Tortosa, de ahí la diferencia en los p -obtenidos (Véase Figura B.26b y Figura 3.4b). Todas las desviaciones en el estudio de ambas colas se producen antes en el tiempo, intensificándose en las últimas décadas.

Ubicación Estadísticos	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa	Madrid
Cola superior, U	2010				1985	1985	1990
Cola inferior, L	2015	1995	2015	1950	2010	1920	1960
Ambas colas, D	2010	1995	1945	1950	1985	1915	1955

Cuadro 3.7: Fechas inicio desviaciones en temperaturas mínimas por localidad.

3.3.3. Análisis por estaciones

Como se pudo ver para Madrid, otra hipótesis de interés es analizar el cambio climático por estaciones del año. A continuación, se tratarán las dos estaciones que inicialmente puedan crear una mayor curiosidad para el lector, verano e invierno. Para ambas estaciones se mostrarán los p -valores obtenidos para el estudio de ambas colas en las series de temperaturas mínimas y máximas. Las tablas con información sobre el resto de estadísticos y estaciones se encuentran en el Apéndice B.

Localidad	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa	Madrid
T_x	< 2.2e-16	< 2.2e-16	0.105	0.004	0.003	6e-04	2e-04
T_m	0.075	0.002	0.013	0.082	2e-04	2e-04	< 2.2e-16

Cuadro 3.8: p -valores obtenidos en el estudio de ambas colas en verano para temperaturas máximas (T_x) y mínimas (T_m) por localidad.

La Tabla 3.8 refleja como en todas ubicaciones, salvo en Cáceres, se detectan desviaciones de estacionariedad en la estación de verano para las temperaturas máximas. Por el contrario, para temperaturas

mínimas, las localidades de Castellón y Huelva presentan p -valores en la frontera. Destacar Daroca y Castellón, donde las evidencias de desviaciones son muy significativas para temperaturas máximas. Comparando ambos tipos de temperatura se puede concluir la existencia de evidencias en contra de la estacionariedad en la estación de verano.

Por otro lado, la Tabla 3.9, refleja los p -valores obtenidos para los análisis llevados a cabo en las series en invierno. En comparación con verano, las evidencias de desviaciones no son tan notables. En el caso de temperaturas máximas, Tortosa y Castellón son las únicas que presentan evidencias significativas. En el caso de temperaturas mínimas, las evidencias parecen más destacables en Izaña, Tortosa y Madrid. En general, al contrario que verano, dependiendo de la ciudad y tipo de temperatura que se estudia se puede observar como la estación de invierno no tiene porque presentar evidencias en contra de estacionariedad en ambas colas. Dichos resultados pueden ser apoyados por las gráficas de número de récords del Apéndice B.

Localidad	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa	Madrid
T_x	0.004	0.051	0.07	0.182	0.107	4e-04	0.045
T_m	0.107	0.517	0.097	0.132	0.02	2e-04	0.006

Cuadro 3.9: p -valores obtenidos en el estudio de ambas colas en invierno para temperaturas máximas (T_x) y mínimas (T_m) por localidad.

3.4. Conclusiones

En el contexto de calentamiento global, es claro el interés de estudiar la existencia de comportamiento no estacionario en las colas de una serie, en especial la de sus récords. En este trabajo hemos revisado el *Modelo Clásico de Récords* y se ha propuesto herramientas estadísticas, basadas en las propiedades de la ocurrencia de récords de series i.i.d., para detectar este tipo de comportamiento.

En particular, en este trabajo, se han estudiado series centenarias de temperaturas máximas y mínimas diarias en 7 localidades de España, Madrid, Castellón, Daroca, Tortosa, Huelva, Cáceres e Izaña (Tenerife). A través de un análisis formal gráfico y numérico se han obtenido fuertes evidencias, salvo para Cáceres, a favor de un comportamiento no estacionario en las colas de la distribución que afecta a la ocurrencia de récords. Tanto para temperaturas máximas como mínimas, para la mayoría de localidades, dicho comportamiento comienza a tomar relevancia entre 1980 y 2000, siendo este muy notable para las localidades de Izaña y Madrid. Cabe destacar el comportamiento de la serie de temperaturas mínimas de Tortosa, la cual presenta desviaciones claras respecto al comportamiento estacionario a partir de 1915 para el estudio de la cola inferior y ambas colas.

Por último se ha realizado un estudio por estaciones con el objetivo de observar una posible homogeneidad. En el caso de Madrid, dicho análisis ha demostrado un comportamiento no estacionario más débil en la estación de Invierno. Para el resto de localidades se concluye que no existe homogeneidad entre las diferentes estaciones del año. La estación que, en general, presenta más evidencias a favor del efecto del calentamiento global es la estación de verano para temperaturas máximas.

Apéndice A

Análisis descriptivo previo

Tenemos series de temperaturas maximas y minimas. El formato en el cual tenemos los datos muestran una columna como solo fecha, entonces vamos a dividir esta en tres columnas, Year, Month y Day. Esto nos podrá hacer de ayuda en según que cálculos.

Contamos con 8 estaciones donde se han hecho mediciones. Estas son Castellón, Madrid-Retiro, Daroca, Ciudad Real, Caceres, Huelva, Izaña y Tortosa.

Todas series, salvo la de Tortosa van desde el 1 de enero de 1920 al 31 de enero de 2024. La de Tortosa cuenta con más datos, desde el 1 de enero de 1905 al 31 de enero de 2024. Para el estudio de análisis eliminaremos los datos correspondientes a la fecha 29 de febrero ya que esta aparece cada 4 años y puede afectar a los resultados del trabajo.

Vamos a estudiar casos en los que los datos pueden presentar problemas para su análisis.

A.1. Datos incorrectos

Un caso interesante es el caso en que la temperatura mínima medida en una fecha supere a la temperatura máxima. En nuestro caso ninguna estación muestra datos incorrectos de este tipo.

A.2. Datos perdidos

Otra cosa importante a estudiar es el numero de NA's que existen en nuestros datos y como están distribuidos. Miraremos primero el total y después por décadas. En este apartado trabajaré con los datos modificados, es decir con NA's en aquellos que hemos considerado incorrectos.

A.2.1. Total NA's

Estación	Total NA's
Castellon	397
Madrid Retiro	175
Daroca	186
Ciudad Real	938
Caceres A	101
Huelva	158
Izaña	413
Tortosa	318

Cuadro A.1: Total NA's por estación.

Como vemos, el numero de datos perdidos es relativamente pequeño en todas estaciones salvo en Ciudad Real. Veamos si es necesario acortar dicha serie en esa estación.

A.2.2. NA's por década

Por antigüedad de datos, uno puede pensar que la mayoría de datos perdidos deberían encontrarse en las primeras décadas.

Esta hipótesis se cumple en todas estaciones.

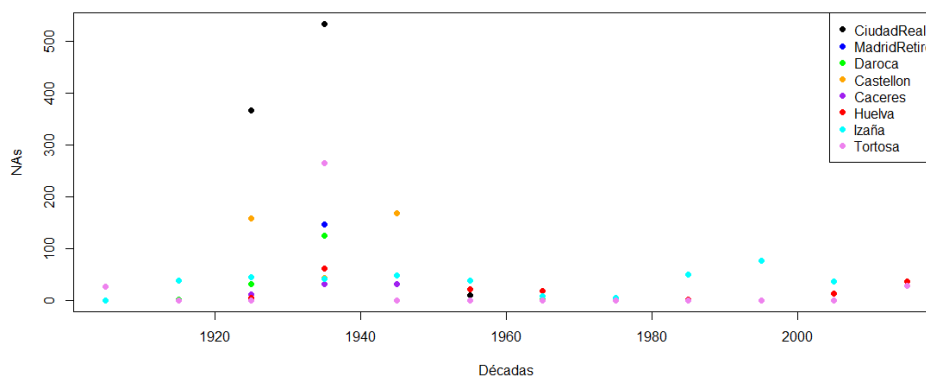


Figura A.1: NAs por década

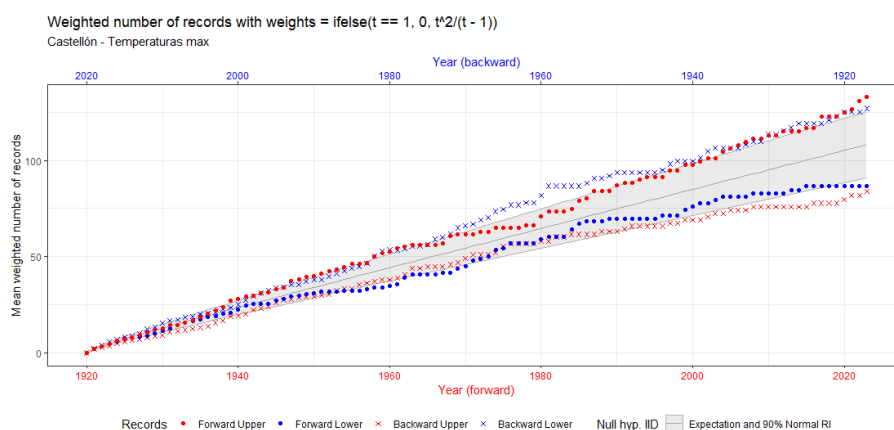
Se observa que para Ciudad Real y Tortosa el número de NA's es mayor en la década de 1935. Por la localización y cantidad de datos perdidos, se elimina la serie de datos correspondiente a Ciudad Real.

Apéndice B

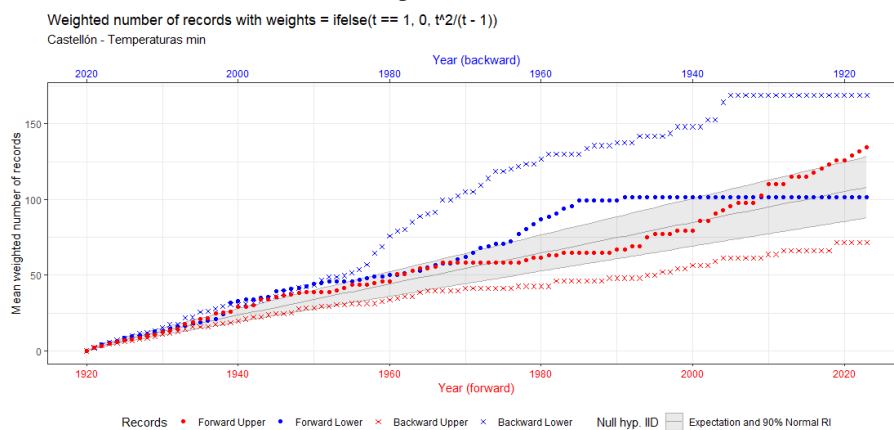
Figuras y tablas

Anexo complementario con las correspondientes gráficas realizadas con los datos de Castellón, Daroca, Tortosa, Huelva, Cáceres e Izaña, y tablas resumen de los p -valores obtenidos por estación.

B.1. Castellón

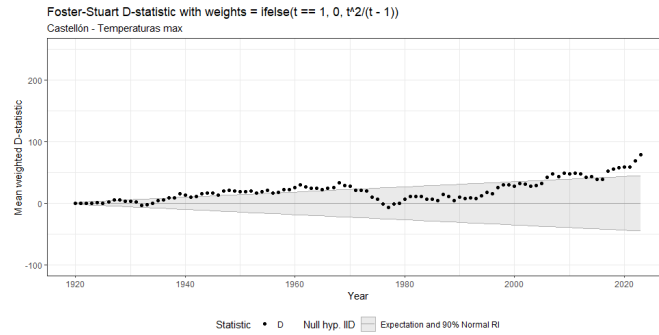


(a) Temperaturas máximas

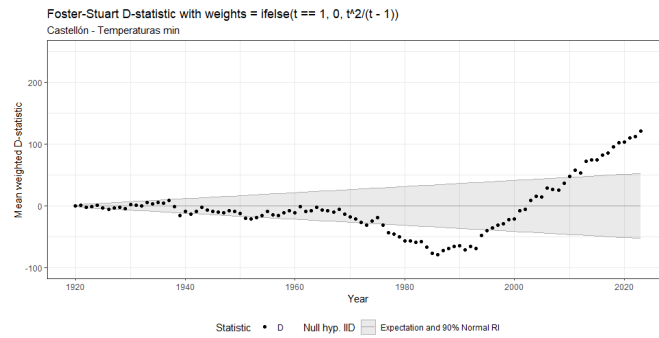


(b) Temperaturas mínimas

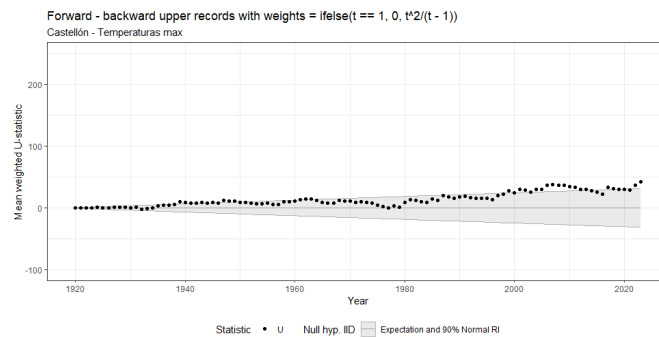
Figura B.1: Media número de récords para temperaturas máximas y mínimas.



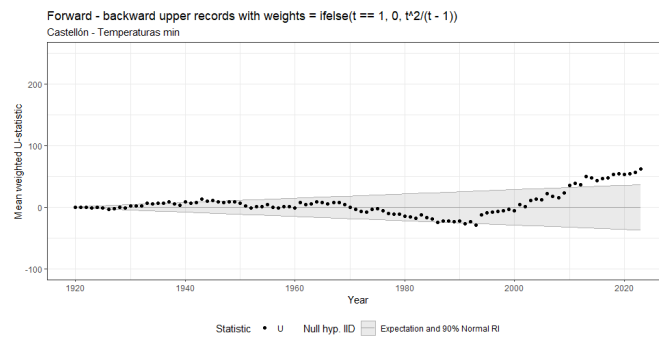
(a) Temperaturas máximas



(b) Temperaturas mínimas

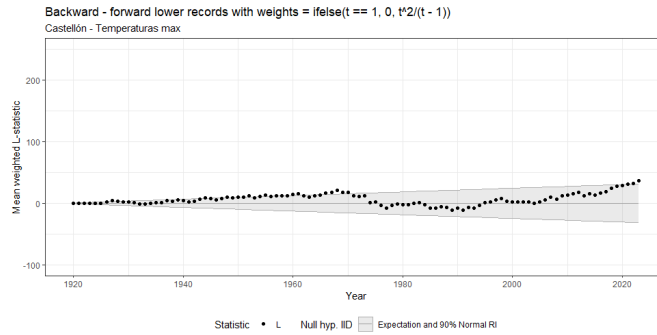
Figura B.2: Valor medio estadístico D para temperaturas máximas y mínimas.

(a) Temperaturas máximas

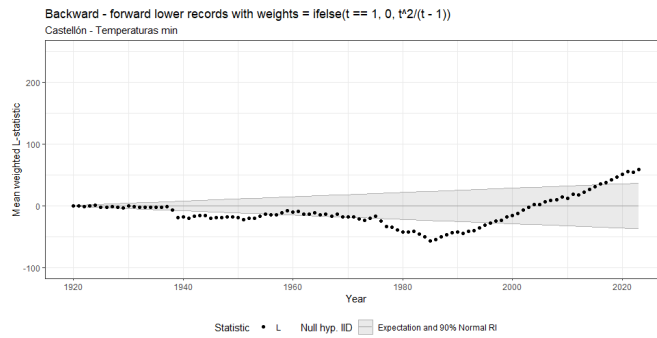


(b) Temperaturas mínimas

Figura B.3: Valor medio estadístico U para temperaturas máximas y mínimas.



(a) Temperaturas máximas



(b) Temperaturas mínimas

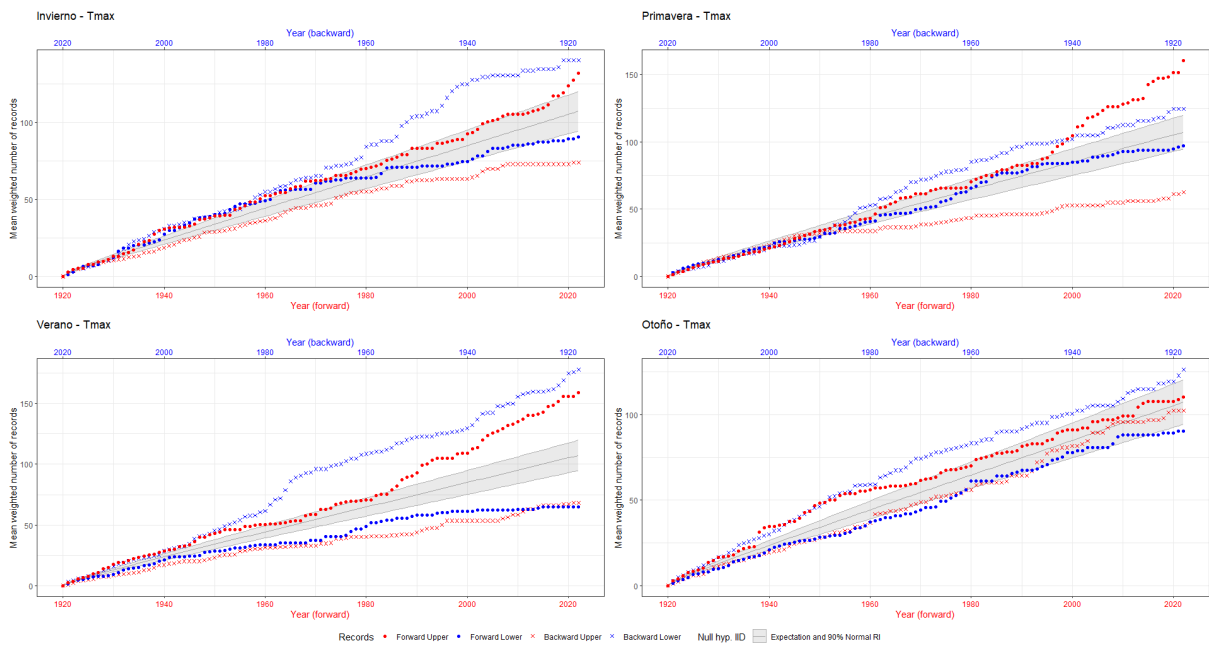
Figura B.4: Valor medio estadístico L para temperaturas máximas y mínimas.

Figura B.5: Media número de récords para temperaturas máximas por estación.

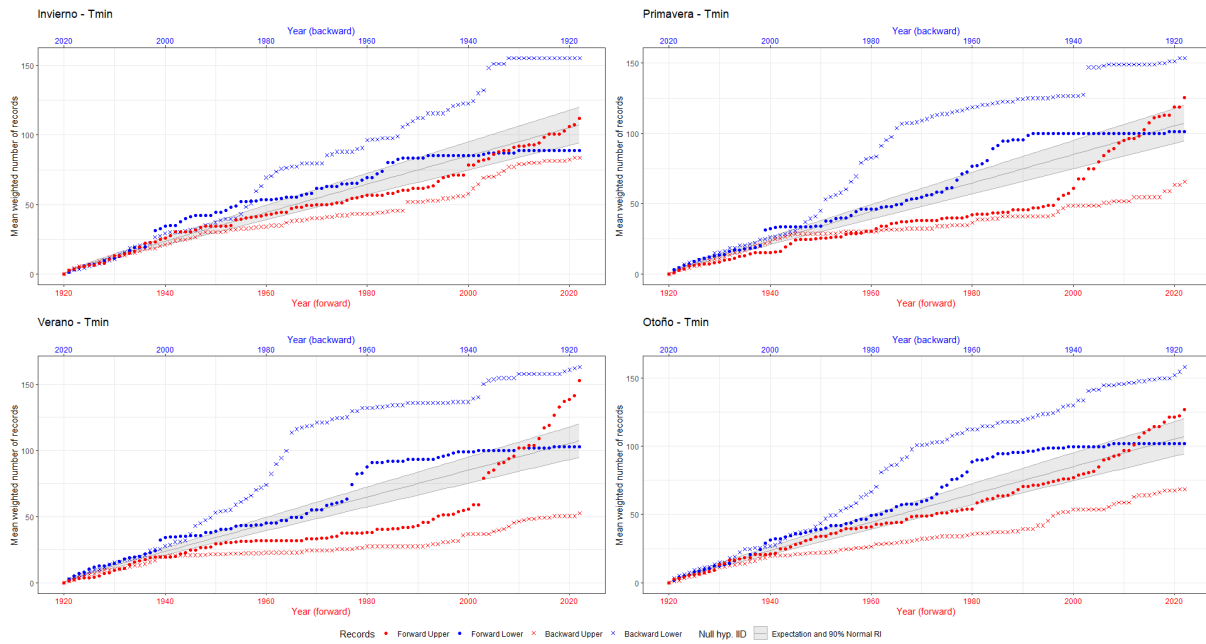
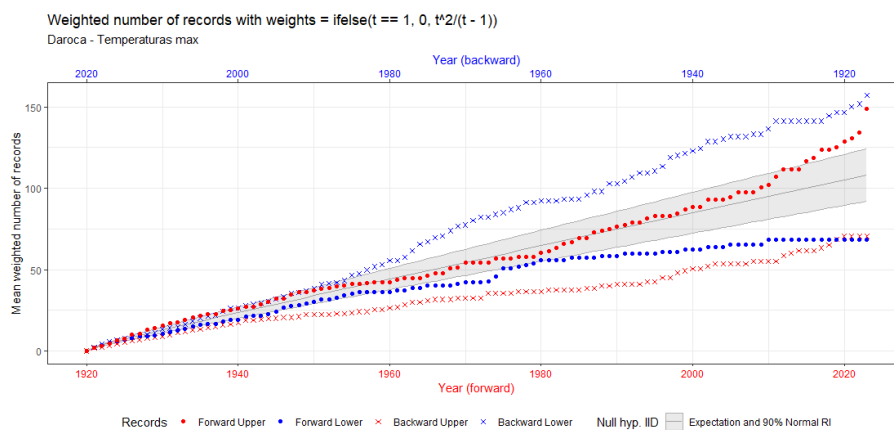
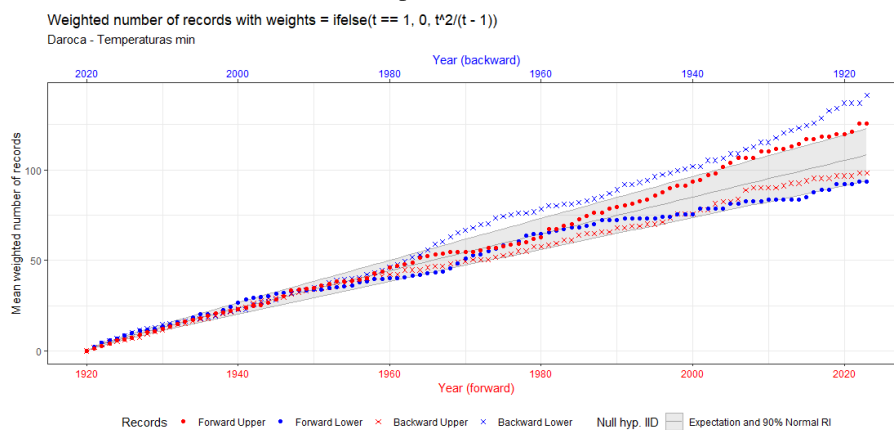


Figura B.6: Media número de récords para temperaturas mínimas por estación.

B.2. Daroca

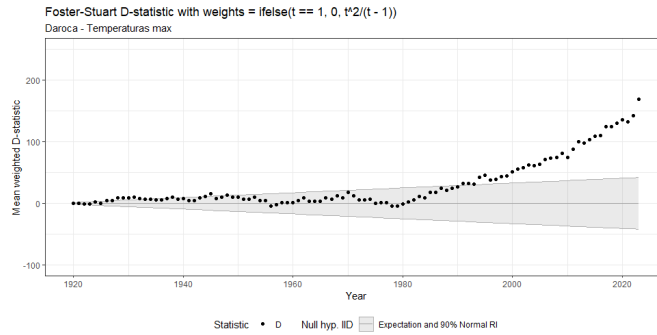


(a) Temperaturas máximas

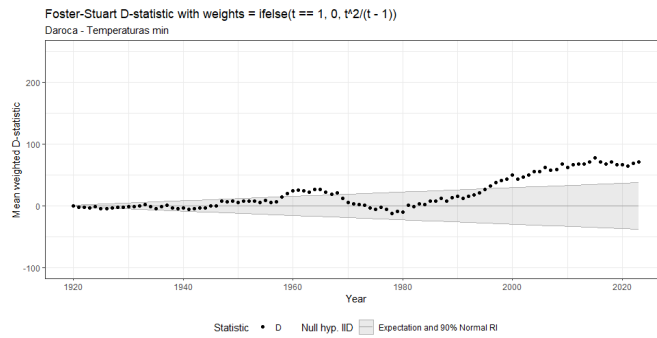


(b) Temperaturas mínimas

Figura B.7: Media número de récords para temperaturas máximas y mínimas.

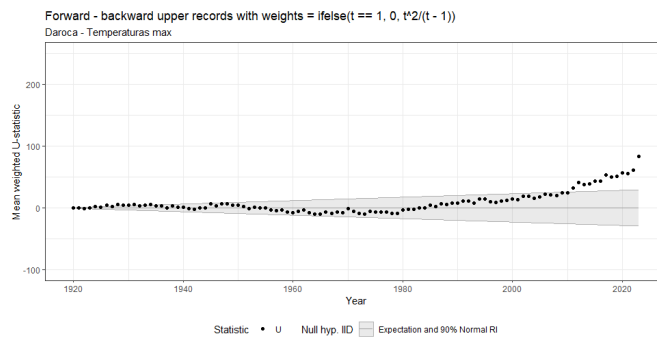


(a) Temperaturas máximas

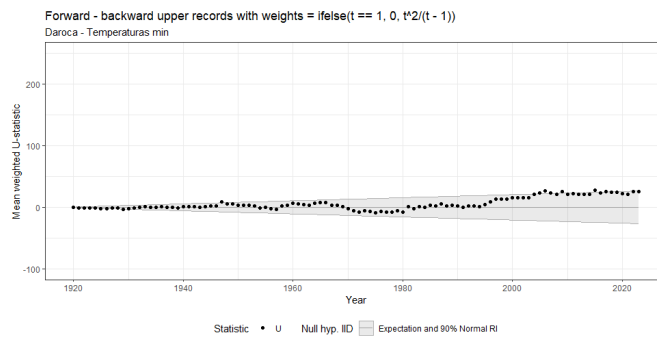


(b) Temperaturas mínimas

Figura B.8: Valor medio estadístico D para temperaturas máximas y mínimas.

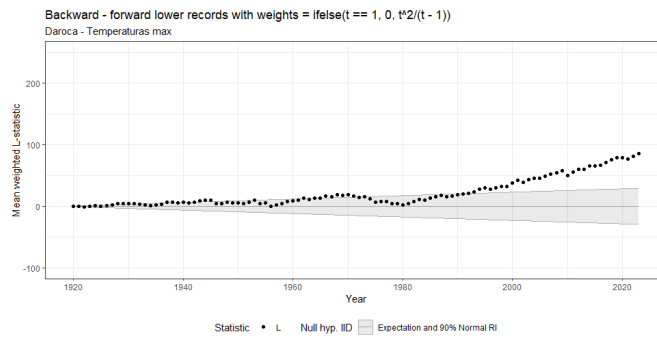


(a) Temperaturas máximas

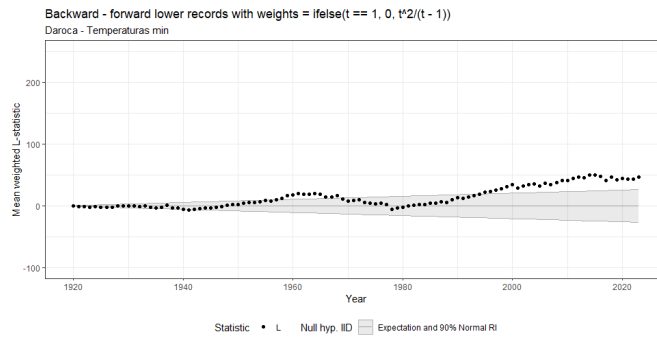


(b) Temperaturas mínimas

Figura B.9: Valor medio estadístico U para temperaturas máximas y mínimas.



(a) Temperaturas máximas



(b) Temperaturas mínimas

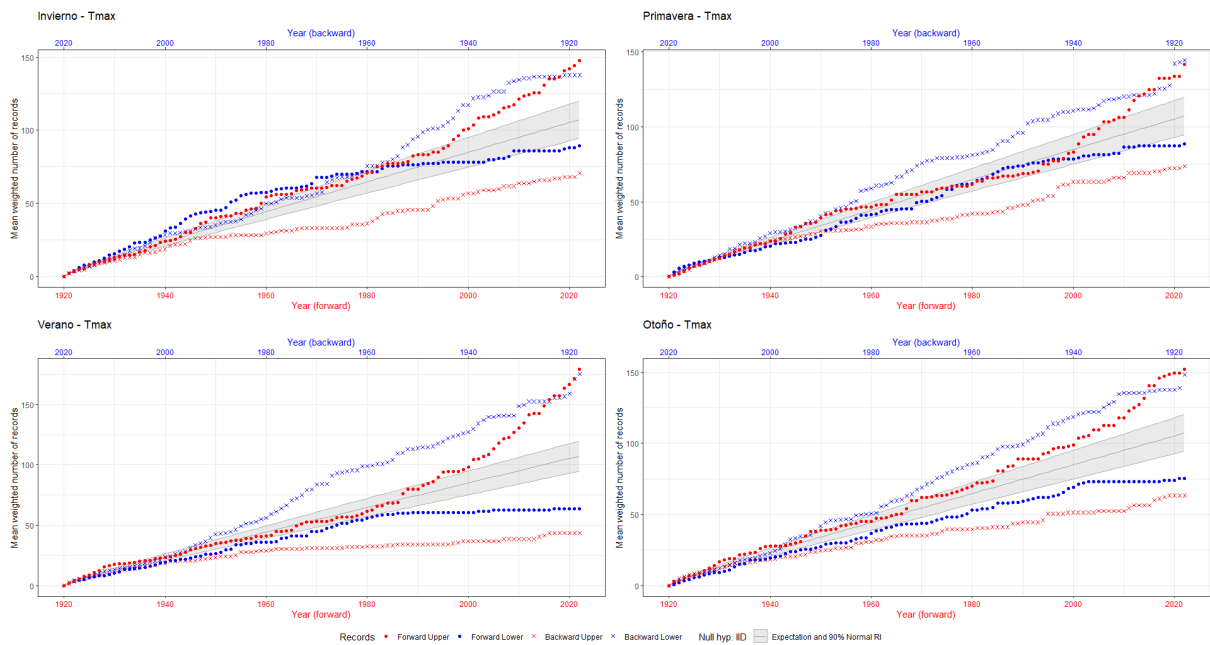
Figura B.10: Valor medio estadístico L para temperaturas máximas y mínimas.

Figura B.11: Media número de récords para temperaturas máximas por estación.

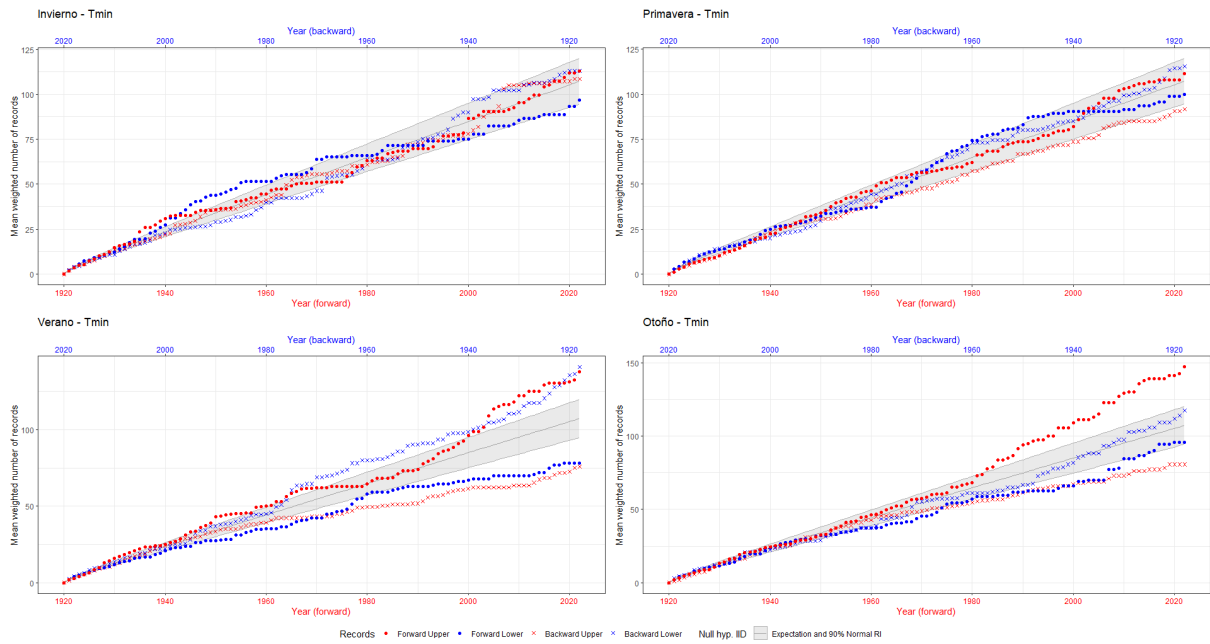
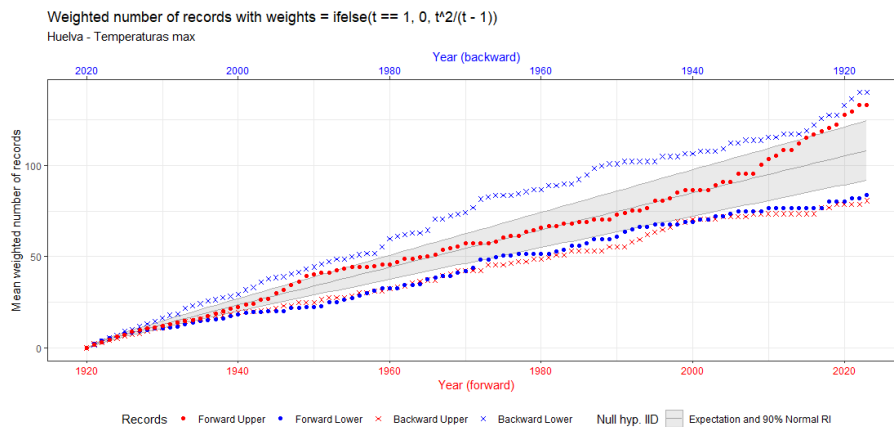
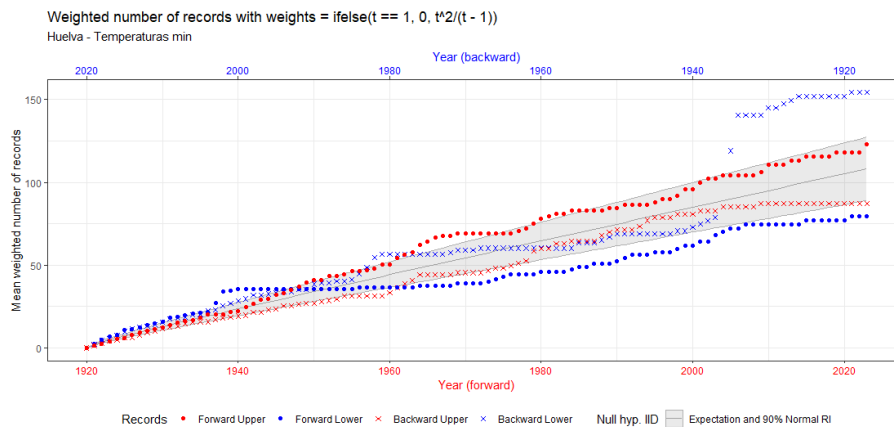


Figura B.12: Media número de récords para temperaturas mínimas por estación.

B.3. Huelva

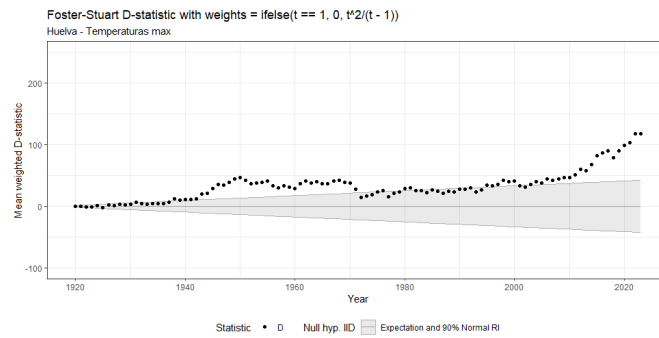


(a) Temperaturas máximas

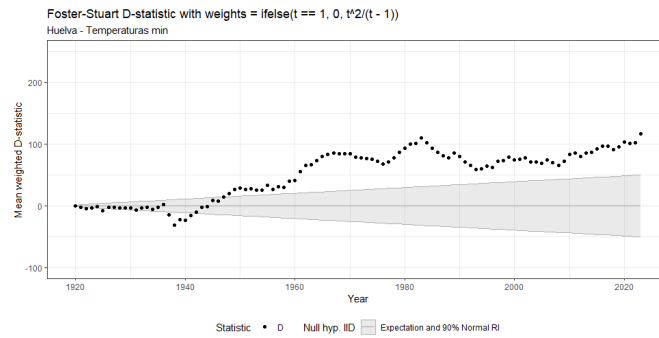


(b) Temperaturas mínimas

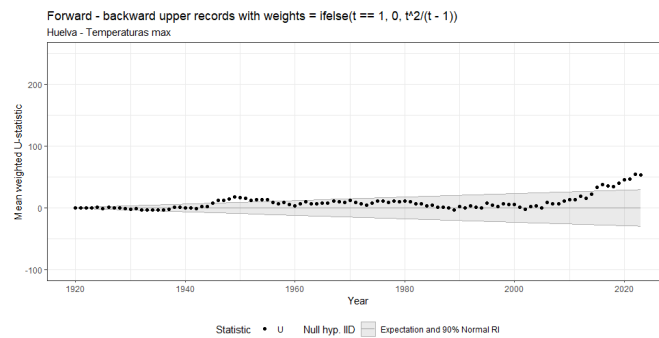
Figura B.13: Media número de récords para temperaturas máximas y mínimas.



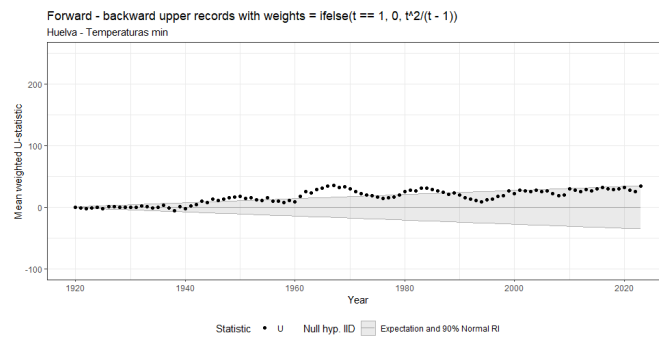
(a) Temperaturas máximas



(b) Temperaturas mínimas

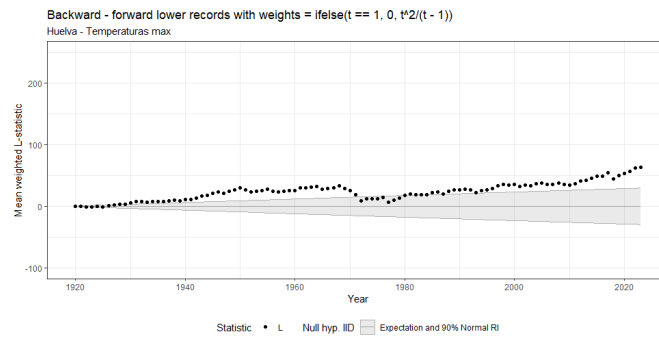
Figura B.14: Valor medio estadístico D para temperaturas máximas y mínimas.

(a) Temperaturas máximas

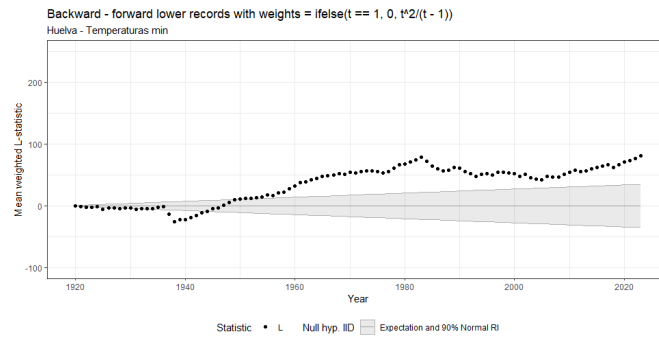


(b) Temperaturas mínimas

Figura B.15: Valor medio estadístico U para temperaturas máximas y mínimas.



(a) Temperaturas máximas



(b) Temperaturas mínimas

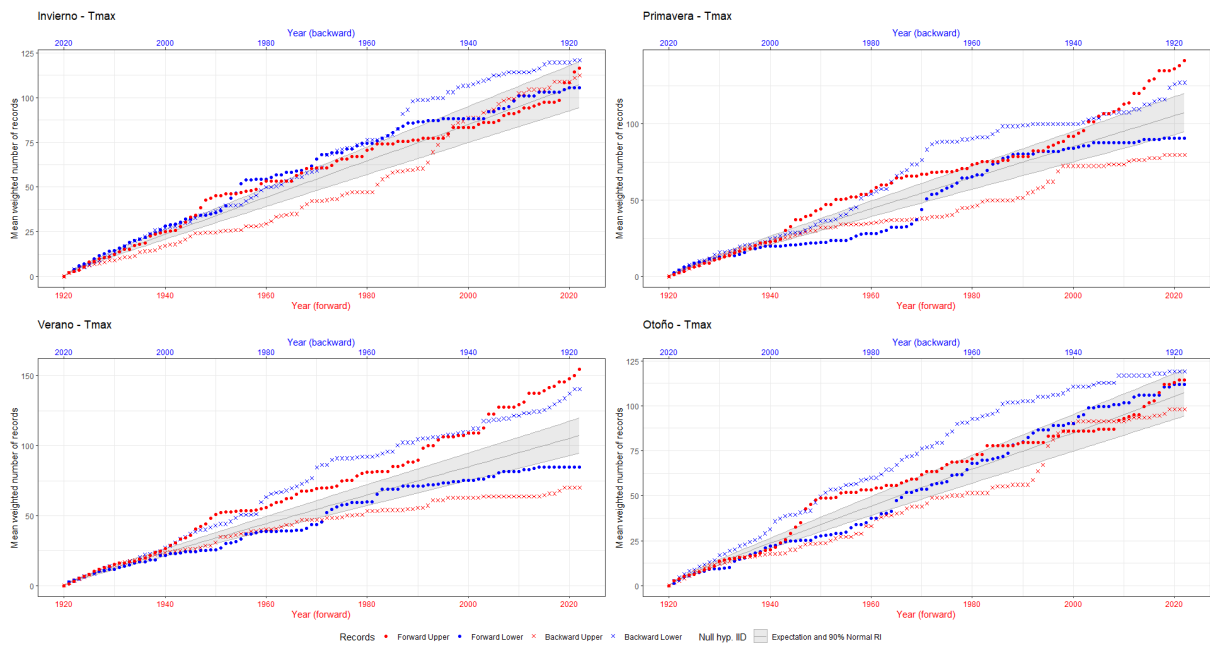
Figura B.16: Valor medio estadístico L para temperaturas máximas y mínimas.

Figura B.17: Media número de récords para temperaturas máximas por estación.

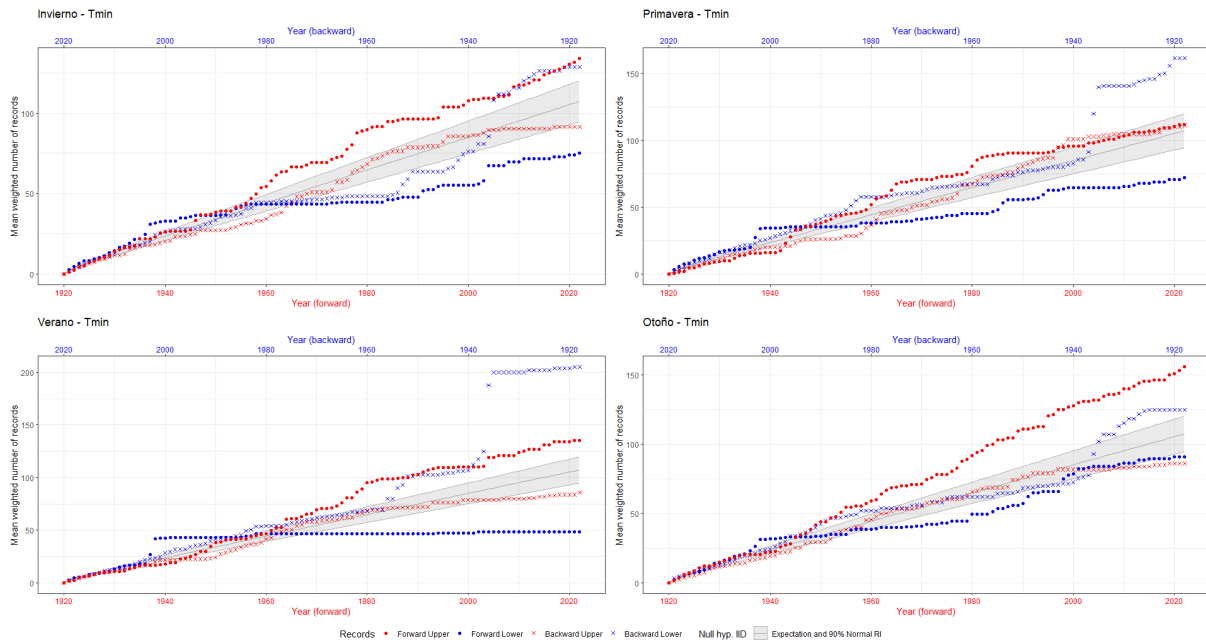
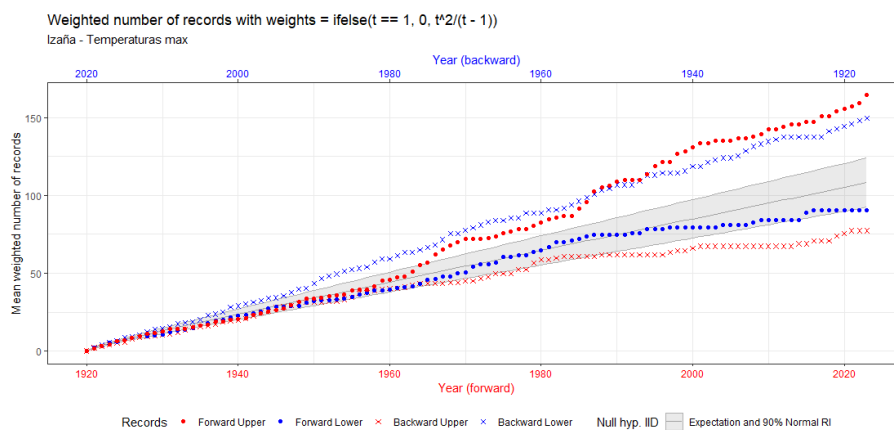
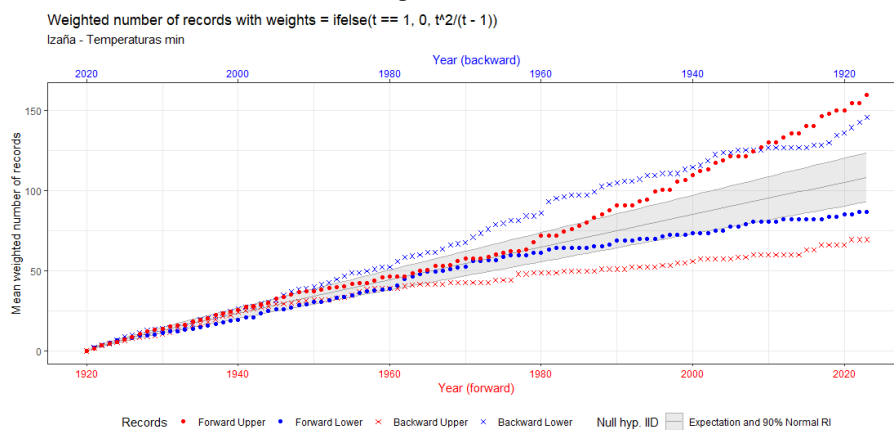


Figura B.18: Media número de récords para temperaturas mínimas por estación.

B.4. Izaña

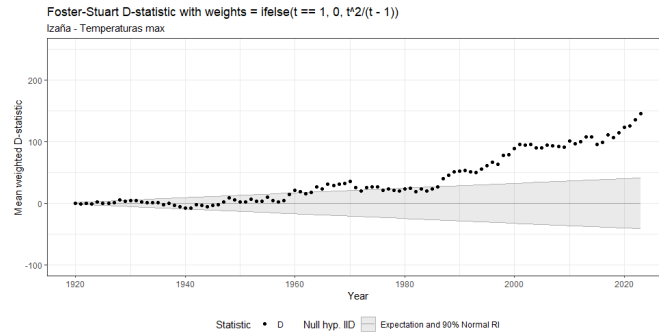


(a) Temperaturas máximas

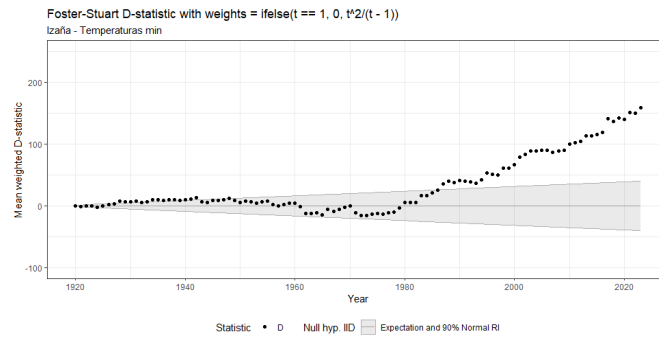


(b) Temperaturas mínimas

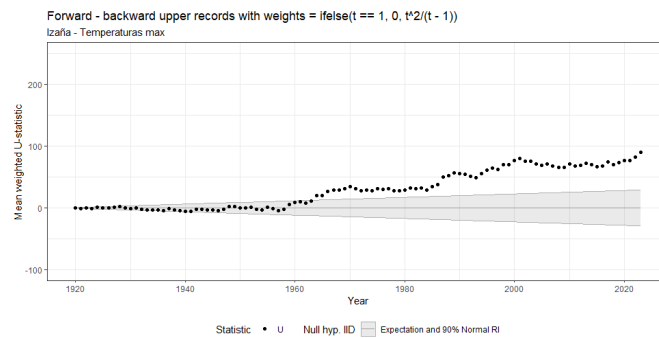
Figura B.19: Media número de récords para temperaturas máximas y mínimas.



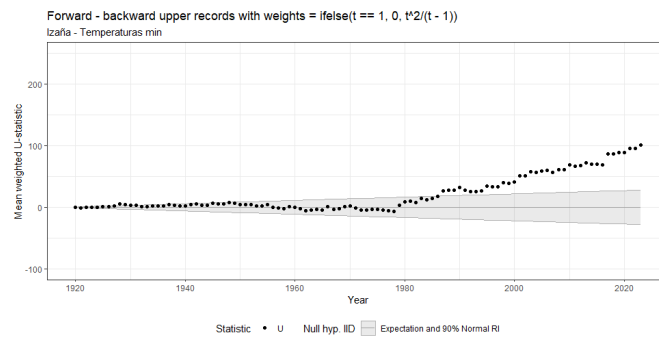
(a) Temperaturas máximas



(b) Temperaturas mínimas

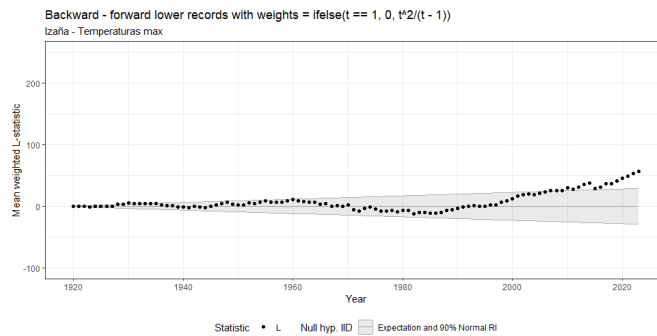
Figura B.20: Valor medio estadístico D para temperaturas máximas y mínimas.

(a) Temperaturas máximas

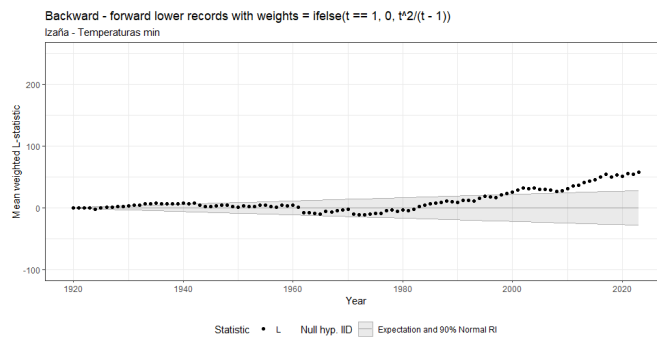


(b) Temperaturas mínimas

Figura B.21: Valor medio estadístico U para temperaturas máximas y mínimas.



(a) Temperaturas máximas



(b) Temperaturas mínimas

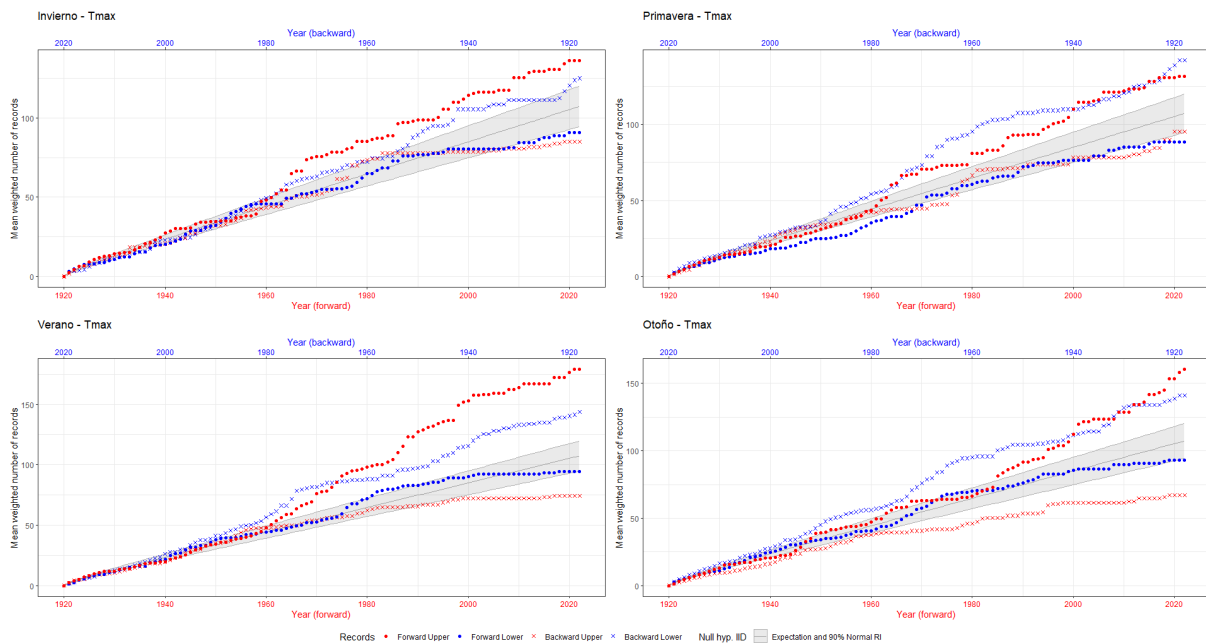
Figura B.22: Valor medio estadístico L para temperaturas máximas y mínimas.

Figura B.23: Media número de récords para temperaturas máximas por estación.

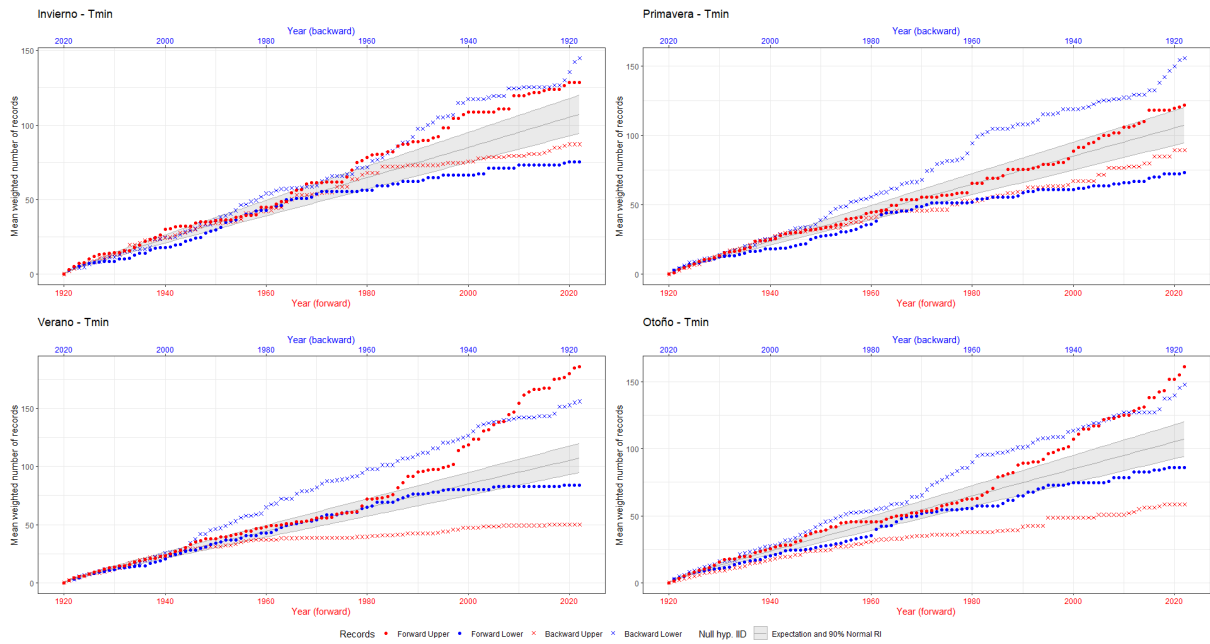
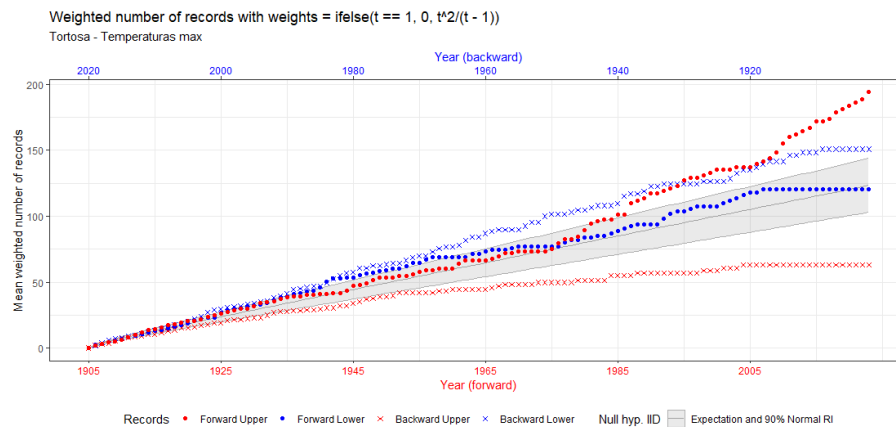
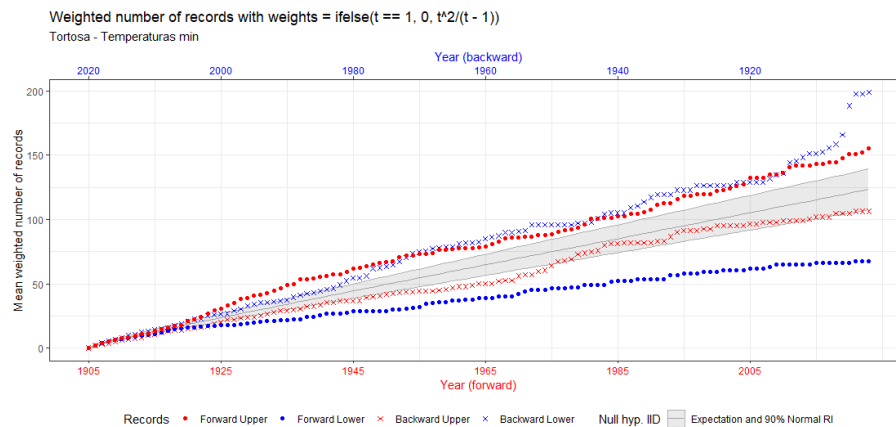


Figura B.24: Media número de récords para temperaturas mínimas por estación.

B.5. Tortosa

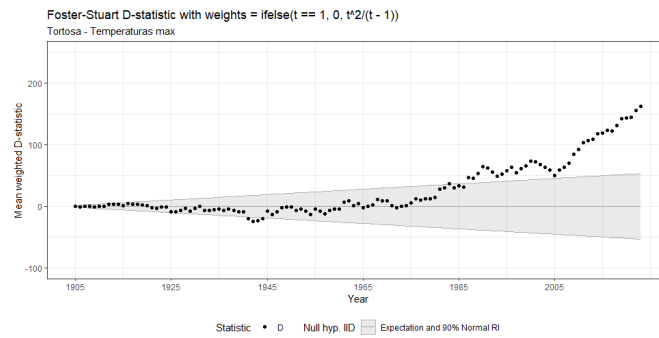


(a) Temperaturas máximas

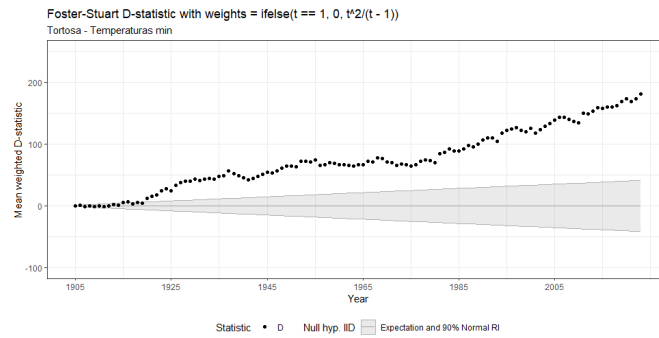


(b) Temperaturas mínimas

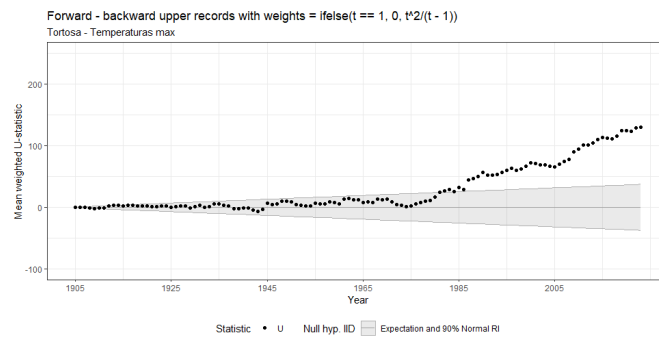
Figura B.25: Media número de récords para temperaturas máximas y mínimas.



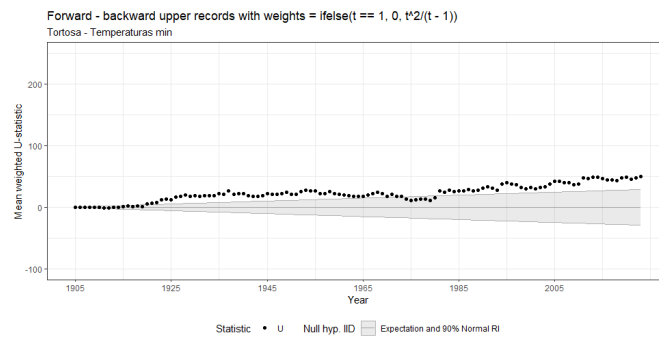
(a) Temperaturas máximas



(b) Temperaturas mínimas

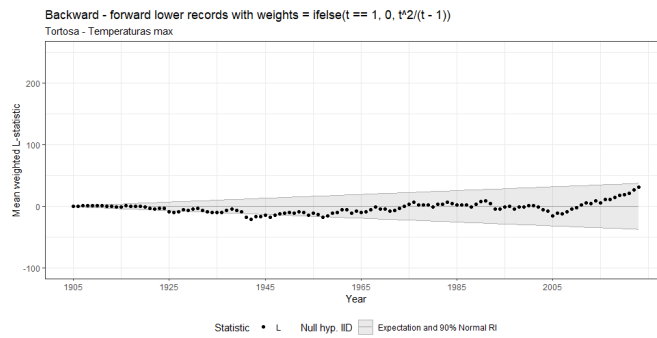
Figura B.26: Valor medio estadístico D para temperaturas máximas y mínimas.

(a) Temperaturas máximas

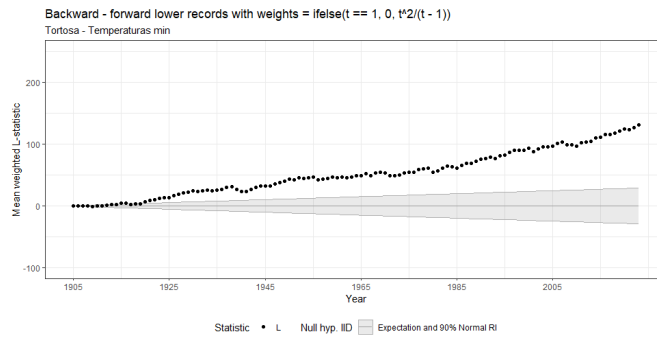


(b) Temperaturas mínimas

Figura B.27: Valor medio estadístico U para temperaturas máximas y mínimas.



(a) Temperaturas máximas



(b) Temperaturas mínimas

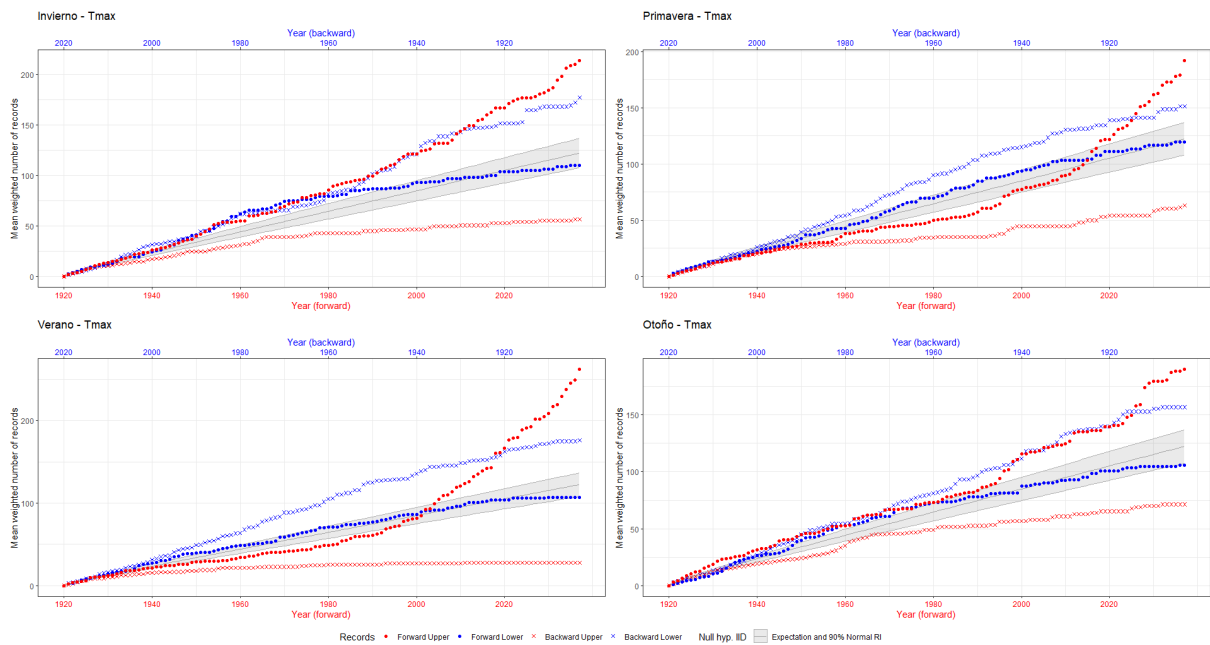
Figura B.28: Valor medio estadístico L para temperaturas máximas y mínimas.

Figura B.29: Media número de récords para temperaturas máximas por estación.

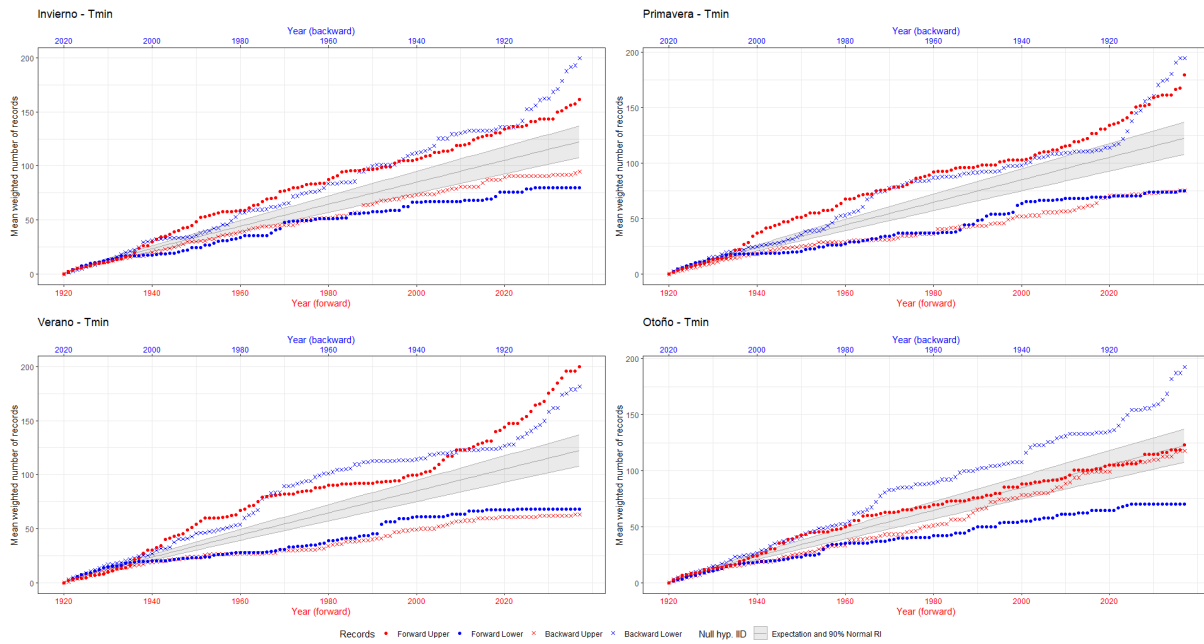
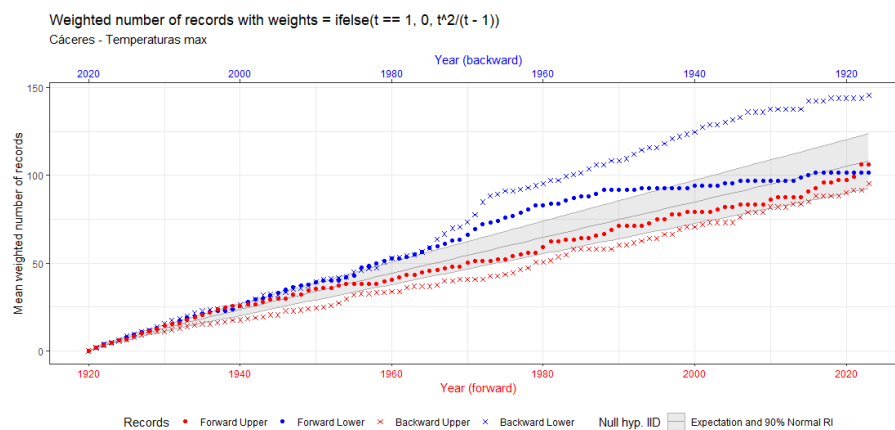
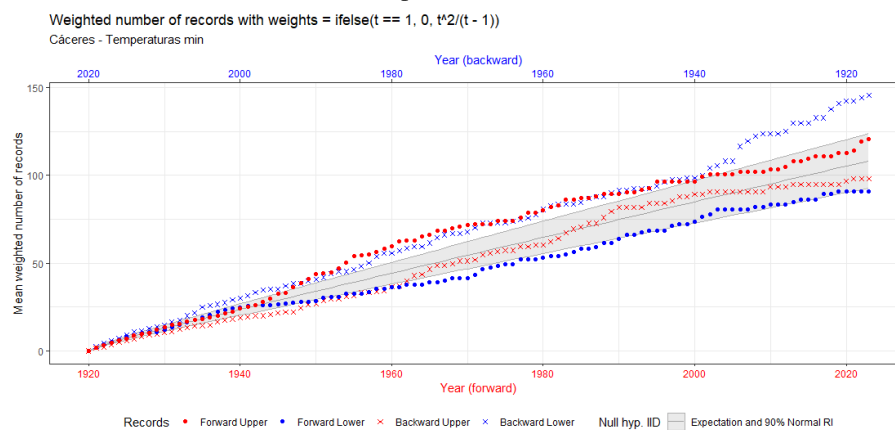


Figura B.30: Media número de récords para temperaturas mínimas por estación.

B.6. Cáceres

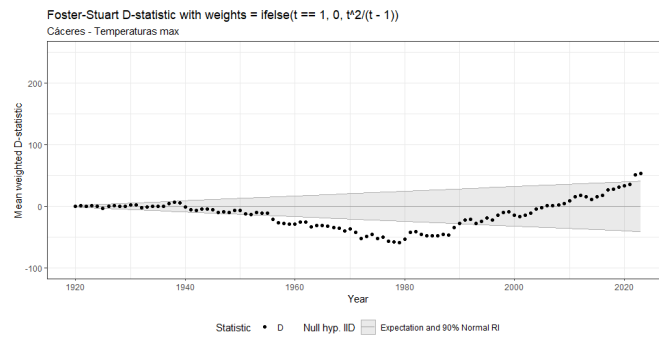


(a) Temperaturas máximas

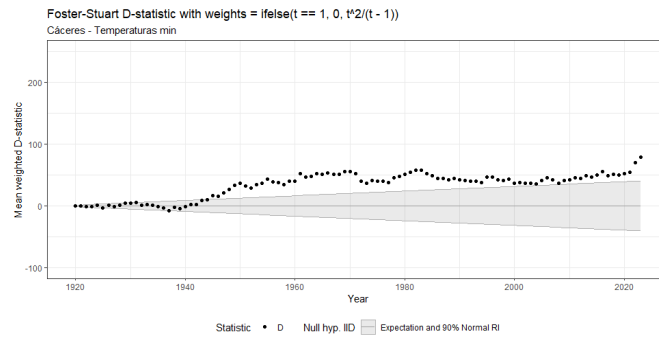


(b) Temperaturas mínimas

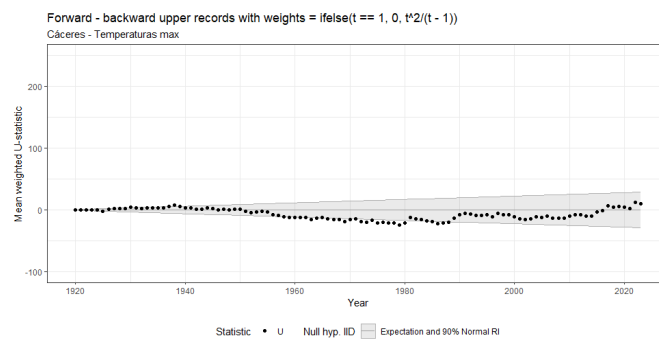
Figura B.31: Media número de récords para temperaturas máximas y mínimas.



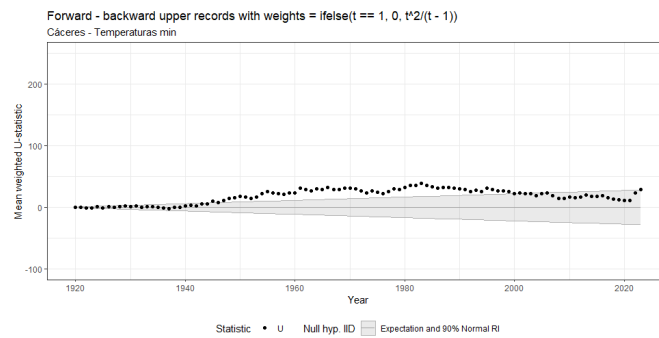
(a) Temperaturas máximas



(b) Temperaturas mínimas

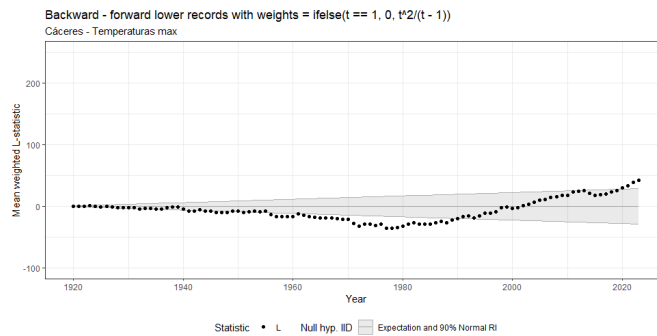
Figura B.32: Valor medio estadístico D para temperaturas máximas y mínimas.

(a) Temperaturas máximas

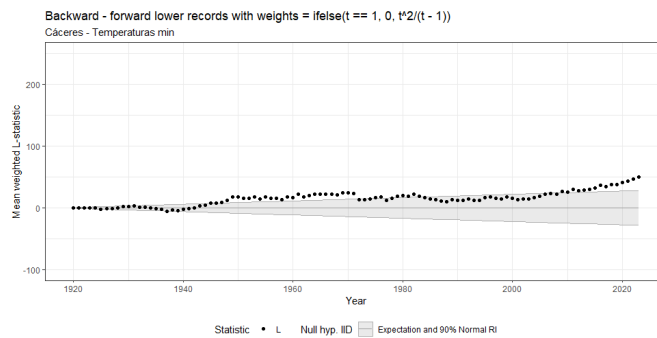


(b) Temperaturas mínimas

Figura B.33: Valor medio estadístico U para temperaturas máximas y mínimas.



(a) Temperaturas máximas



(b) Temperaturas mínimas

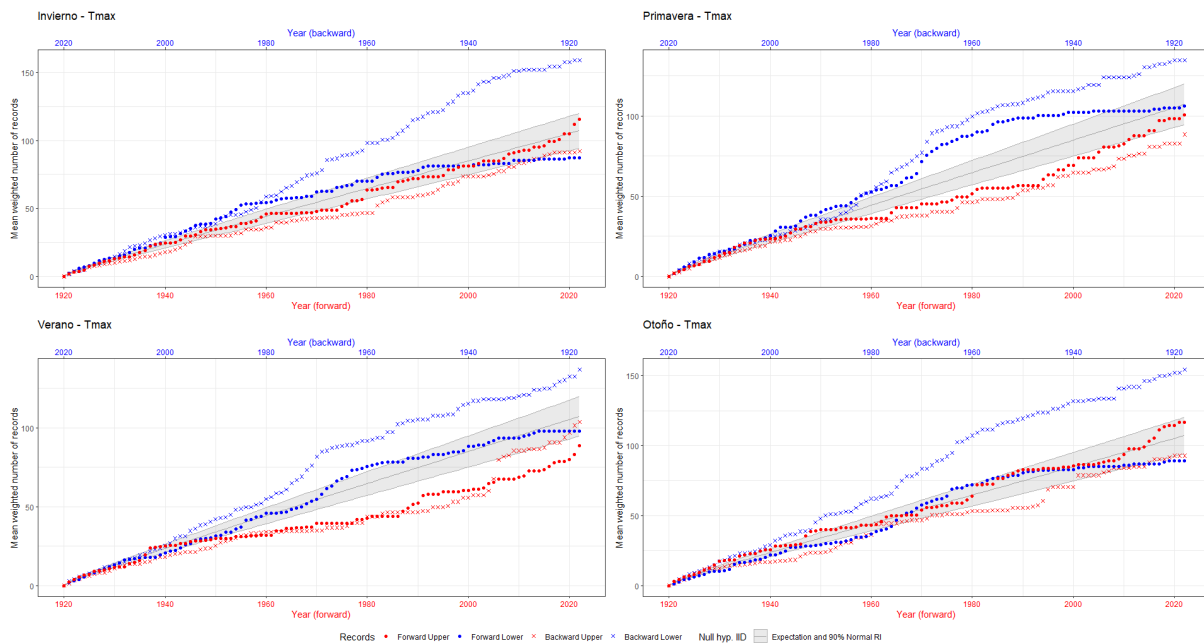
Figura B.34: Valor medio estadístico L para temperaturas máximas y mínimas.

Figura B.35: Media número de récords para temperaturas máximas por estación.

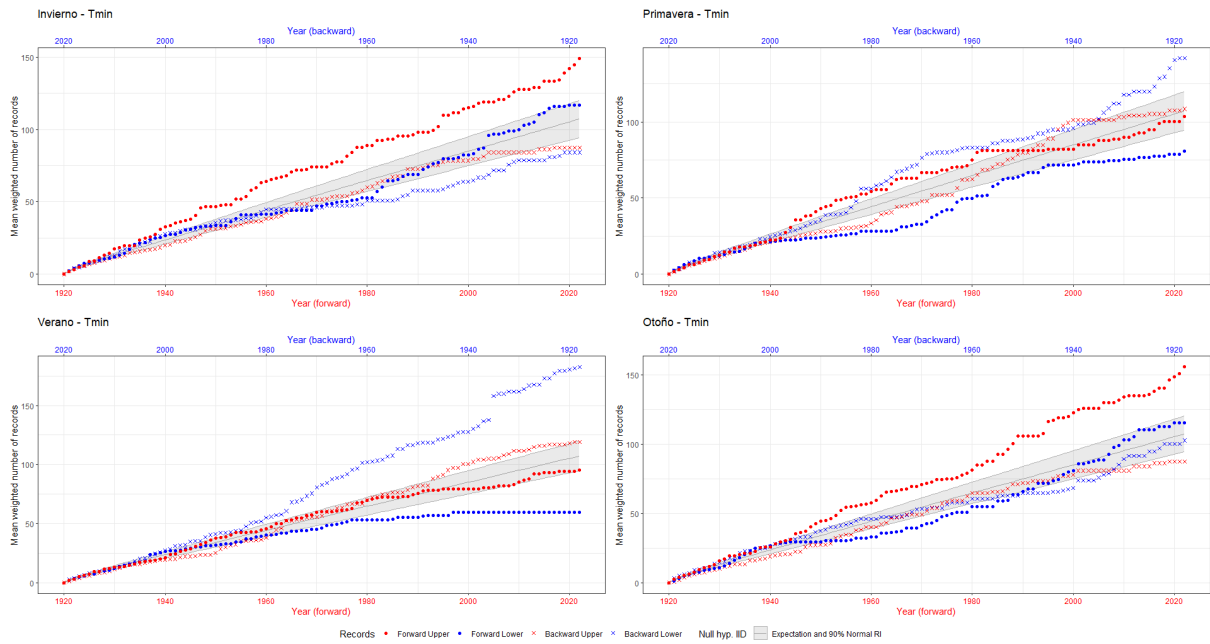


Figura B.36: Media número de récords para temperaturas mínimas por estación.

B.7. Madrid (resto)

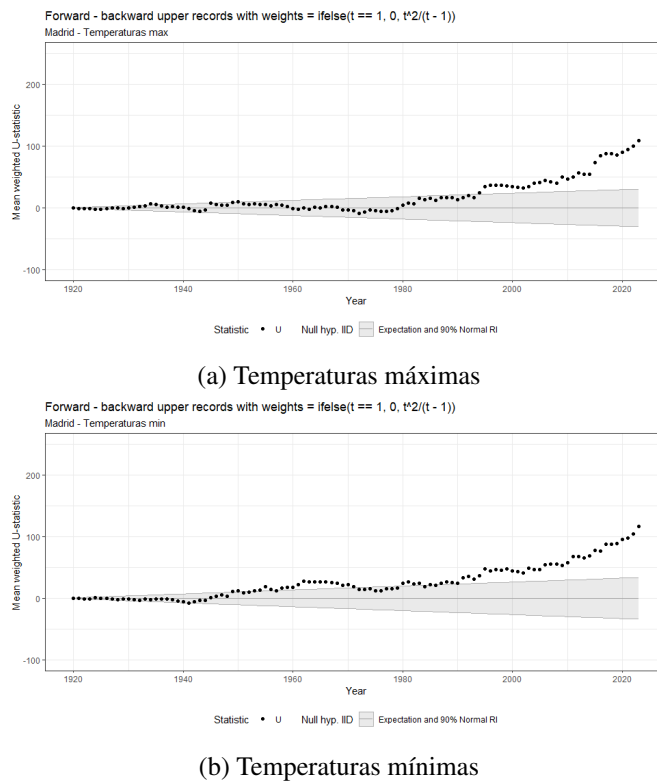


Figura B.37: Valor medio estadístico U para temperaturas máximas y mínimas.

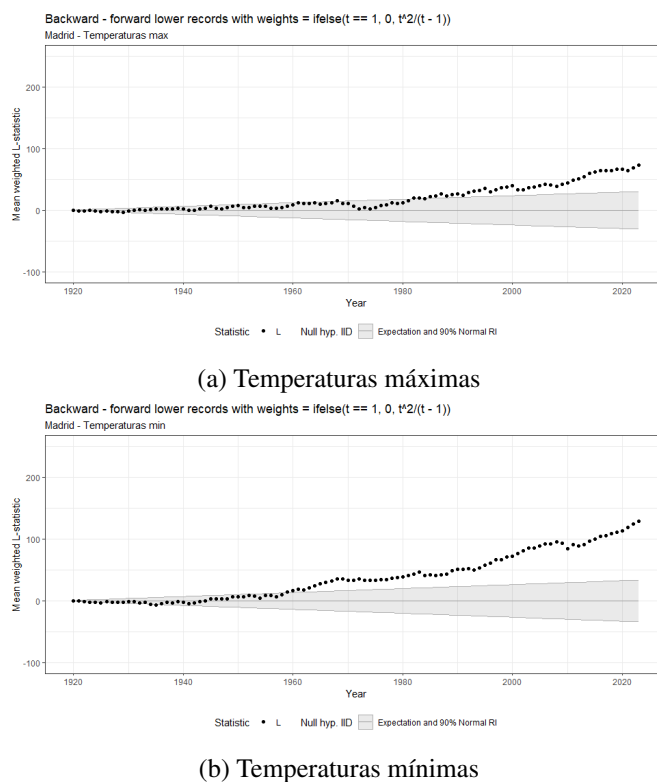


Figura B.38: Valor medio estadístico L para temperaturas máximas y mínimas.

B.8. Mapas

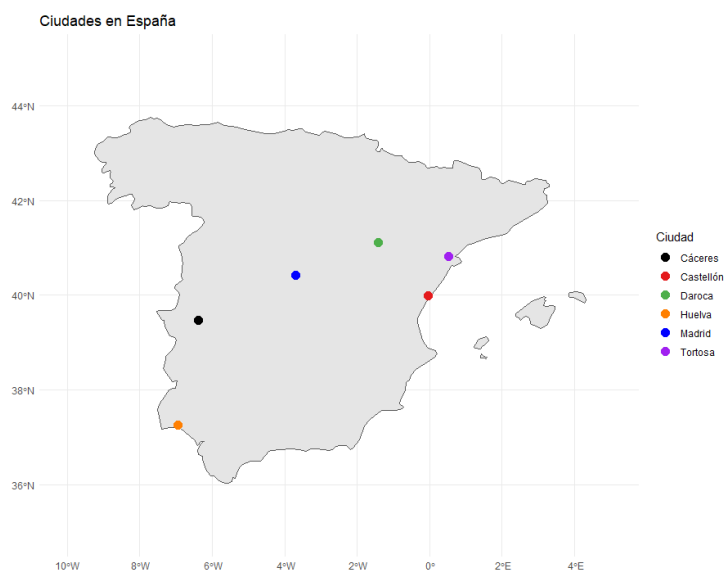


Figura B.39: Localidades a estudiar de la Península Ibérica

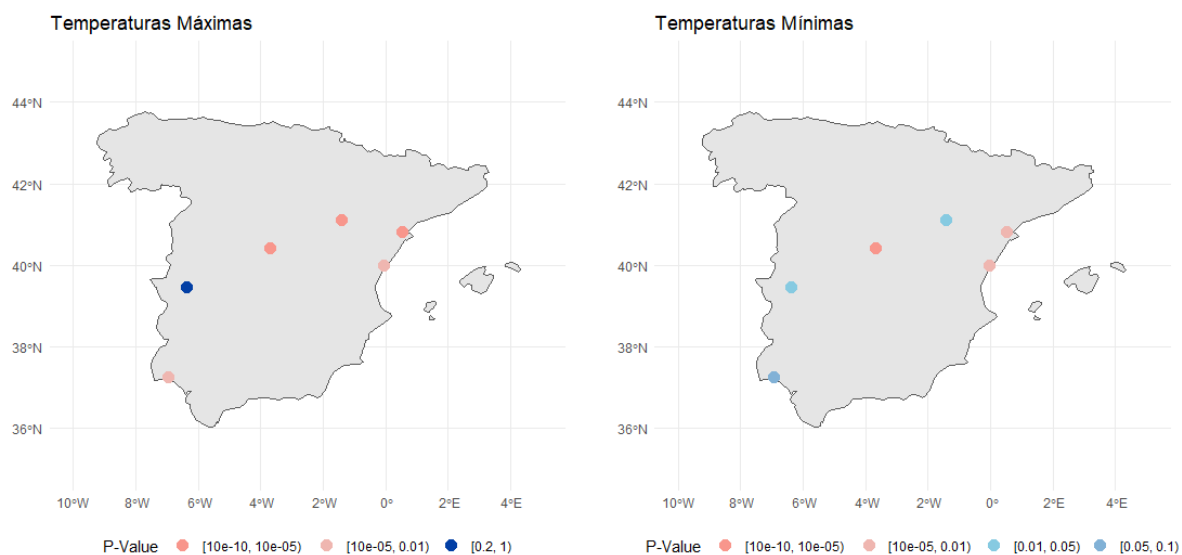


Figura B.40: P-valores del análisis del efecto de una tendencia positiva en la cola superior para las localidades de Daroca, Castellón, Huelva, Cáceres, Madrid y Tortosa.

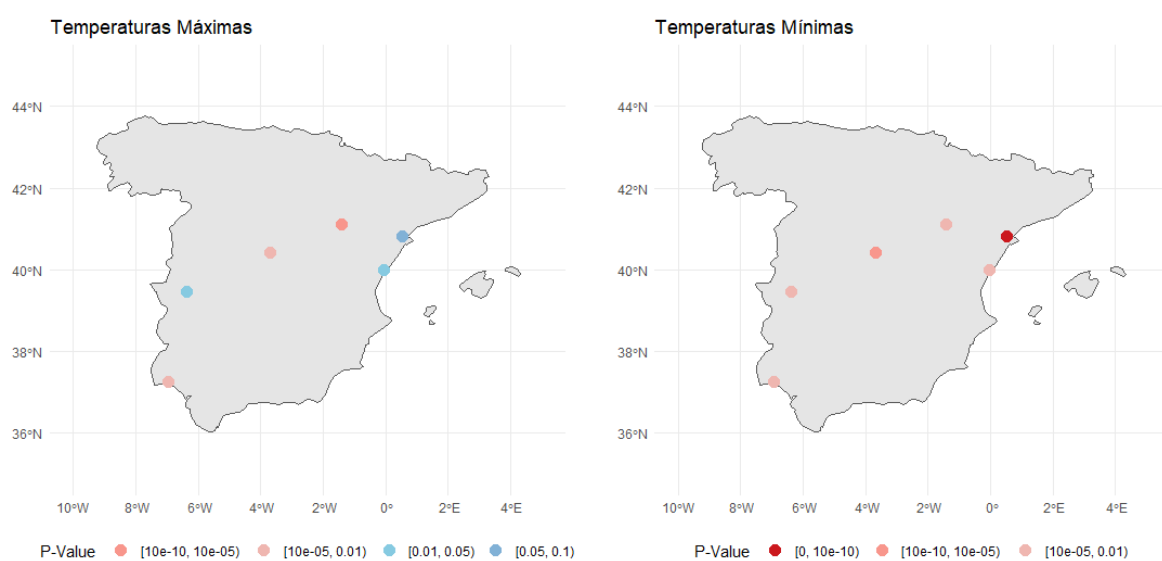


Figura B.41: P-valores del análisis del efecto de una tendencia positiva en la cola inferior para las localidades de Daroca, Castellón, Huelva, Cáceres, Madrid y Tortosa.

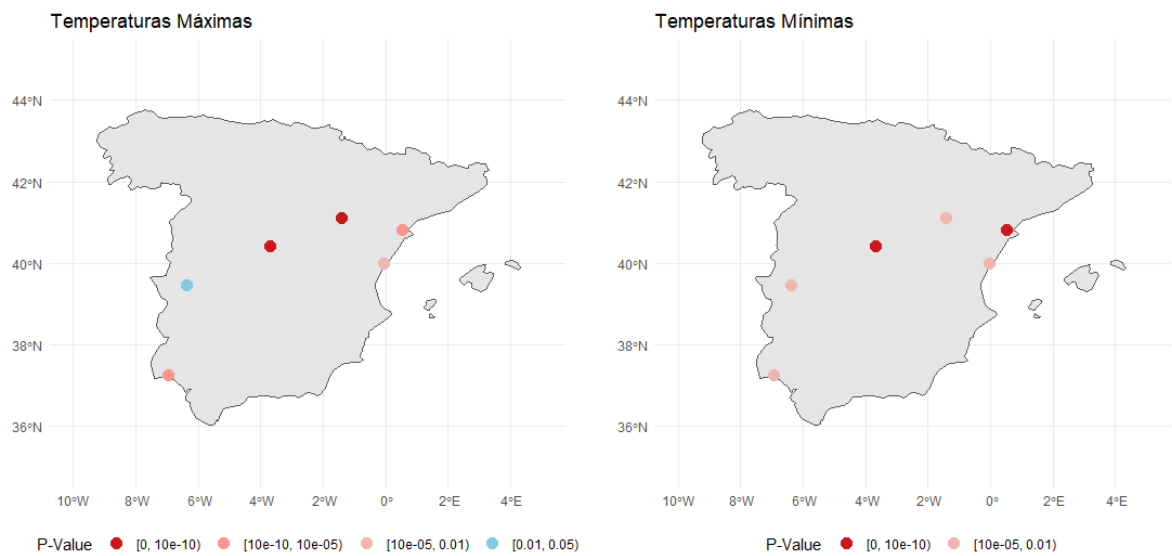


Figura B.42: P-valores del análisis del efecto de una tendencia positiva en ambas colas para las localidades de Daroca, Castellón, Huelva, Cáceres, Madrid y Tortosa.

B.9. Tablas por estación

Ubicación Estadísticos	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa
Cola superior, U	2e-04	<2.2e-16	0.345	0.004	0.003	2e-04
Cola inferior, L	<2.2e-16	<2.2e-16	0.024	0.014	0.048	0.011
Ambas colas, D	<2.2e-16	<2.2e-16	0.105	0.004	0.003	6e-04

Cuadro B.1: p -valores de distintos estadísticos de temperaturas máximas diarias en verano por ubicación.

Ubicación Estadísticos	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa
Cola superior, U	0.115	0.018	0.431	0.136	8e-04	2e-04
Cola inferior, L	0.141	0.012	4e-04	0.087	0.001	0.001
Ambas colas, D	0.075	0.002	0.0.13	0.082	2e-04	2e-04

Cuadro B.2: p -valores de distintos estadísticos de temperaturas mínimas diarias en verano por ubicación.

Ubicación Estadísticos	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa
Cola superior, U	0.006	0.003	0.168	0.117	0.104	< 2.2e-16
Cola inferior, L	0.061	0.453	0.046	0.461	0.293	0.113
Ambas colas, D	0.004	0.051	0.07	0.182	0.107	4e-04

Cuadro B.3: p -valores de distintos estadísticos de temperaturas máximas diarias en invierno por ubicación.

Ubicación Estadísticos	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa
Cola superior, U	0.197	0.302	7e-04	0.04	0.061	8e-04
Cola inferior, L	0.137	0.666	0.713	0.361	0.049	0.005
Ambas colas, D	0.107	0.517	0.097	0.1324	0.02	2e-04

Cuadro B.4: p -valores de distintos estadísticos de temperaturas mínimas diarias en invierno por ubicación.

Ubicación Estadísticos	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa
Cola superior, U	0.045	0.033	0.273	0.043	0.141	0.02
Cola inferior, L	0.331	0.03	0.276	0.056	0.0032	0.125
Ambas colas, D	0.0924	0.014	0.261	0.032	0.011	0.029

Cuadro B.5: p -valores de distintos estadísticos de temperaturas máximas diarias en primavera por ubicación.

Ubicación Estadísticos	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa
Cola superior, U	0.2379	0.285	0.223	0.44	0.154	$< 2.2\text{e-}16$
Cola inferior, L	0.084	0.335	0.014	0.11	0.002	0.001
Ambas colas, D	0.112	0.31	0.052	0.203	0.011	$< 2.2\text{e-}16$

Cuadro B.6: p -valores de distintos estadísticos de temperaturas mínimas diarias en primavera por ubicación.

Ubicación Estadísticos	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa
Cola superior, U	0.016	6e-04	0.074	0.138	4e-04	$< 2.2\text{e-}16$
Cola inferior, L	2e-04	2e-04	0.00586	6e-04	0.002	0.022
Ambas colas, D	8e-04	$< 2.2\text{e-}16$	0.0139	0.005	6e-04	$< 2.2\text{e-}16$

Cuadro B.7: p -valores de distintos estadísticos de temperaturas máximas diarias en otoño por ubicación.

Ubicación Estadísticos	Castellón	Daroca	Cáceres	Huelva	Izaña	Tortosa
Cola superior, U	0.021	0.017	2e-04	0.001	$< 2.2\text{e-}16$	0.03
Cola inferior, L	0.022	0.241	0.168	0.191	0.001	2e-04
Ambas colas, D	0.015	0.047	0.013	0.026	$< 2.2\text{e-}16$	$< 2.2\text{e-}16$

Cuadro B.8: p -valores de distintos estadísticos de temperaturas mínimas diarias en otoño por ubicación.

Código R

Análisis descriptivo previo

```
library (lubridate)

#separamos fechas en tres columnas Year, Month and Day (ejmplo tmax)
tortosa_tmax$Year<-year(tortosa_tmax$DATE)
tortosa_tmax$Month<-month(tortosa_tmax$DATE)
tortosa_tmax$Day<-day(tortosa_tmax$DATE)
tmax$Year<-year(tmax$DATE)
tmax$Month<-month(tmax$DATE)
tmax$Day<-day(tmax$DATE)

#eliminar datos 29 febrero
ind<-which(tmax$Month==2 & tmax$Day==29)
tmax<-tmax[-ind,]

ind2<-which(tortosa_tmin$Month==2 & tortosa_tmin$Day==29)
tortosa_tmax<-tortosa_tmax[-ind2,]

#datos incorrectos tmax<tmin
incorrectos<- function(data1,data2,columna){
  ind<-which(data1[[columna]]<data2[[columna]])
  return(ind)
}

#ejemplo
incor<-incorrectos(tmax,tmin,2)

#separo por decadas por ejemplo
# Obtener el año mínimo y máximo
# Obtener el año mínimo y máximo (ejemplos)
min_year <- floor(min(tmax$Year))
max_year <- ceiling(max(tmax$Year))

x<-c(1,2,4,5,6,10,12,42,51)
breaks<-c(0,10,20,30,40,50)
cut(x,breaks)

breaks<-seq(min_year-1, max_year+10, by = 10)
labels<-seq(min_year,max_year,by=10)
tmax$Decada<-cut(tmax$Year,breaks,labels)

##NAS (decada)
estaciones<-c(0,'Castellon','MAdridRetiro','Daroca','CiudadReal','Caceres','Huelva',
              'Izana','Tortosa')
```



```

cont_NA_dec<-function(columna){
  count<-numeric(length(labels2))
  for (i in 0:length(count)){
    count[i]<-sum(is.na(subset(tmin[[columna]],tmin$Decada==labels2[i])))
  }
  #grafica
  plot(labels2, count, xlab = "Décadas", ylab = 'Datos incorrecos',
        main = paste('Estacion',estaciones[columna]), ylim = c(0, max(count) + 1),pch=19)

  return(count)
}

cont_NA_dec2<-function(columna){
  count<-numeric(length(labels))
  for (i in 0:length(count)){
    count[i]<-sum(is.na(subset(tortosa_tmin[[columna]],tortosa_tmin$Decada==labels[i])))
  }
  #grafica
  plot(labels, count, xlab = "Décadas", ylab = 'Datos incorrecos',
        main = paste('Estacion',estaciones[columna]), ylim = c(0, max(count) + 1),pch=19)

  return(count)
}

contCast<-cont_NA_dec(2)
contMr<-cont_NA_dec(3)
contDar<-cont_NA_dec(4)
contCr<-cont_NA_dec(5)
contCac<-cont_NA_dec(6)
contHu<-cont_NA_dec(7)
contIz<-cont_NA_dec(8)
contTor<-cont_NA_dec2(2)

#plot NA por decada

xlim <- range(labels)
plot(labels2, contCr, xlab = 'Décadas', ylab = 'NAs', pch = 19, xlim = xlim,
      main='NAs temperatura mínima por década')

points(labels2, contMr, col = "blue", pch = 19)
points(labels2, contDar, col = "green", pch = 19)
points(labels2, contCast, col = "orange", pch = 19)
points(labels2, contCac, col = "purple", pch = 19)
points(labels2, contHu, col = "red", pch = 19)
points(labels2, contIz, col = "cyan", pch = 19)
points(labels, contTor, col='violet', pch=19)

x_coord <- max(labels)
y_coord <- max(contCast, na.rm = TRUE) + 10

legend('topright',legend = c("CiudadReal", "MadridRetiro", "Daroca", "Castellon",
                             "Caceres", "Huelva", "Izaña", 'Tortosa'),
      col = c("black", "blue", "green", "orange", "purple", "red","cyan",'violet'),

```

```
pch = 19)
```

Análisis Preliminar

```
library (RecordTest)
library(dplyr)

#Records y suavizado LOESS (Madrid)
records(tmax$MadridRet)+ggplot2:: geom_smooth(formula = y~x,
                                                method = stats::loess,
                                                mapping = ggplot2::aes(y=tmax$MadridRet),
                                                se = FALSE, col='red')

#obtención periodos 20 años (para tortosa igual)
tmax1<- tmax %>% filter(Year >= 1920 & Year <= 1939)
tmax2 <- tmax %>% filter(Year >= 1960 & Year <= 1979)
tmax3 <- tmax %>% filter(Year >= 2000 & Year <= 2019)

tmin1<- tmin %>% filter(Year >= 1920 & Year <= 1939)
tmin2 <- tmin %>% filter(Year >= 1960 & Year <= 1979)
tmin3 <- tmin %>% filter(Year >= 2000 & Year <= 2019)

#grafico
temp20<-function(df1,df2,df3,col,nombre){
  x1<-series_split(df1[[col]], Mcols=365)
  x2<-series_split(df2[[col]], Mcols=365)
  x3<-series_split(df3[[col]], Mcols=365)

  mx1<-apply(x1,2,mean,na.rm=T)
  mx2<-apply(x2,2,mean,na.rm=T)
  mx3<-apply(x3,2,mean,na.rm=T)

  plot(mx1,type='l',ylim=c(min(mx1,mx2,mx3),max(mx1,mx2,mx3)),
        ,ylab='Temperatura',xlab='Día')
  points(mx2,type='l',col='red')
  points(mx3,type='l',col='blue')

  legend('topright',legend= c('1920-1939', '1960-1979', '2000-2019'),
        col=c('black', 'red', 'blue'), lty = 1)
}

#ejemplo
temp20(tmin1,tmin2,tmin3,3,'Madrid')
```

Análisis datos final

```
library(RecordTest)
library(lubridate)
```

```

library(ggplot2)

#Obtención series incorreladas sin componente estacional
split_uncor<-function(df,col){

  df_split<-series_split(df[[col]],Mcols=365)

  uncor<-series_uncor(df_split,type='adjacent')

  return(uncor)
}

#Deshacer empates
series_untie2<-function(X){

  ncol=ncol(X)
  nrow=nrow(X)
  U<-matrix(runif(nrow*ncol,min=-5*10^(-2),max=5*10^(-2)),nrow=nrow,ncol=ncol)

  return(X+U)
}

#Funcion que muestra todo a utilizar
analisis<-function(df,col,nombre,tipo){

  resultado<-list()

  #Datos y periodo

  resultado$Datos<-paste(nombre,'-',tipo)

  resultado$Periodo<-c(year(min(df[[1]])),year(max(df[[1]])))

  #numero records serie complete (sin transformar)
  rec<-records(df[[col]], plot = FALSE, variable =df[[1]])
  num1<-length(rec$Forward.Upper.record$Times)
  num2<-length(rec$Forward.Lower.record$Times)
  exp<-sum(1/(1:length(df[[col]])))

  records<-data.frame(row.names = 'Num.Records',
                      Upper=num1,
                      Lower=num2,
                      Expected=round(exp,2))

  resultado$num.records<-records

  #separar en M series independientes
  uncor<-split_uncor(df,col)

  #empates
  T1<-series_ties(uncor,record = 'upper')
  T2<-series_ties(uncor, record='lower')

```

```

perc<-data.frame(row.names = 'percentage',
                  Upper=round(T1$percentage,2),Lower=round(T2$percentage,2))

resultado$empates<-perc

uncor2<-series_untie2(uncor)

resultado$Dim<-dim(uncor2)

resultado$Mserie<-uncor2

#numero de records 2 (serie transformada)
X<-uncor2
Xu<-N.test(X)
Xl<-N.test(X,record='lower',alternative='less')

Upper<-Xu$estimate[1]
Lower<-Xl$estimate[1]
Expected<-Xu$estimate[2]

resultado$num.records2<-data.frame(Upper,Lower,Expected)

#tests y p-values U,L y D

D<-foster.test(uncor2,weights=function(t) ifelse(t==1,0,t^2/(t-1)),
               statistic = 'D')
U<-foster.test(uncor2,weights=function(t) ifelse(t==1,0,t^2/(t-1)),
               statistic = 'U')
L<-foster.test(uncor2,weights=function(t) ifelse(t==1,0,t^2/(t-1)),
               statistic = 'L')

pvalues<-data.frame(Statistic=c('U','L','D'),
                    p.values=c(U$p.value,L$p.value,D$p.value))

resultado$p.values<-pvalues

#graficos
data<-resultado$Mserie
subtitle<-paste(nombre,'-',tipo)
breaks<-seq(1, nrow(data), by = 20)
labels<-seq(resultado$Perido[1], 2020, by = 20)

#Escala para gráficos
S1<-ggplot2::scale_x_continuous(breaks = breaks,
                               labels = labels,
                               sec.axis = ggplot2::sec_axis(~ 2021 - .,
                                                            name = "Year (backward)",
                                                            breaks = 1919 + breaks))
S2<-ggplot2::scale_x_continuous(breaks = breaks,
                               labels = labels)
COL<-ggplot2::theme(axis.title.x = ggplot2::element_text(colour = "red"),

```

```

axis.text.x = ggplot2::element_text(colour = "red"),
axis.title.x.top = ggplot2::element_text(colour = "blue"),
axis.text.x.top = ggplot2::element_text(colour = "blue"))

w<-function(t) ifelse(t==1,0,t^2/(t-1))

#Nplot (media número de récords)
g1<-N.plot(data,weights=w)
g1$labels$subtitle<-subtitle
g1$labels$x<-'Year (forward)'
g1<-g1+S1+COL

#fosterplot (valor medio estadístico D)
g2<-foster.plot(data,weights=w,statistic = 'D')
g2$labels$subtitle<-subtitle
g2$labels$x<-'Year'
g2<-g2+S2

print(g1)
print(g2)

return(resultado)
}

#ejemplo
TX<-analisis(tmax,3,'Madrid','Temperaturas max')
TM<-analisis(tmin,3,'Madrid','Temperaturas min')

```

Análisis por estaciones

```

library (RecordTest)
library(dplyr)
library(ggplot2)
library(ggpubr)

#Filtrado por estaciones
winter<-function(df,col){
  #filtrado
  X<-df %>% filter((df[[10]] == 12 & df[[11]] >= 21) |
                  (df[[10]] == 1) |
                  (df[[10]] == 2) |
                  (df[[10]] == 3 & df[[11]] <= 20))

  ind<-which(X[[10]]<4 & X[[9]]==1920)

  X<-X[-ind,]

  ind<-which(X[[10]]==12 & X[[9]]==2023)

  X<-X[-ind,]

```

```

#separacion y empates
Y<-series_split(X[[col]],Mcols=90)
Y<-series_untie2(Y)

return(Y)
}
spring<-function(df,col){
  #filtrado
  X<-df %>% filter((df[[10]] == 3 & df[[11]] >= 22) |
                    (df[[10]] == 4) |
                    (df[[10]] == 5) |
                    (df[[10]] == 6 & df[[11]] <= 21))

  #quitamos primavera ultimo año ya que no tenemos el invierno
  ind<-which(X[[9]]==2023)

  X<-X[-ind,]

  #separacion y empates
  Y<-series_split(X[[col]],Mcols=92)
  Y<-series_untie2(Y)

  return(Y)
}
summer<-function(df,col){
  #filtrado
  X<-df %>% filter((df[[10]] == 6 & df[[11]] >= 22) |
                    (df[[10]] == 7) |
                    (df[[10]] == 8) |
                    (df[[10]] == 9 & df[[11]] <= 23))

  #quitamos primavera ultimo año ya que no tenemos el invierno
  ind<-which(X[[9]]==2023)

  X<-X[-ind,]

  #separacion y empates
  Y<-series_split(X[[col]],Mcols=94)
  Y<-series_untie2(Y)

  return(Y)
}
autumn<-function(df,col){
  #filtrado
  X<-df %>% filter((df[[10]] == 9 & df[[11]] >= 24) |
                    (df[[10]] == 10) |
                    (df[[10]] == 11) |
                    (df[[10]] == 12 & df[[11]] <= 20))

  #quitamos primavera ultimo año ya que no tenemos el invierno
  ind<-which(X[[9]]==2023)

```

```

X<-X[-ind,]

#separacion y empates
Y<-series_split(X[[col]],Mcols=88)
Y<-series_untie2(Y)

return(Y)
}

#tests
pval<-function(df,col){
  pvalues<-list()

  U<-foster.test(winter(df,col),statistic = 'U',permutation.test = T,B=5000)
  L<-foster.test(winter(df,col),statistic = 'L',permutation.test = T,B=5000)
  D<-foster.test(winter(df,col),statistic = 'D',permutation.test = T,B=5000)

  pvalues$winter<-data.frame(Statistic=c('U','L','D'),
                                p.values=c(U$p.value,L$p.value,D$p.value))

  U<-foster.test(spring(df,col),statistic = 'U',permutation.test = T,B=5000)
  L<-foster.test(spring(df,col),statistic = 'L',permutation.test = T,B=5000)
  D<-foster.test(spring(df,col),statistic = 'D',permutation.test = T,B=5000)

  pvalues$spring<-data.frame(Statistic=c('U','L','D'),
                              p.values=c(U$p.value,L$p.value,D$p.value))

  U<-foster.test(summer(df,col),statistic = 'U',permutation.test = T,B=5000)
  L<-foster.test(summer(df,col),statistic = 'L',permutation.test = T,B=5000)
  D<-foster.test(summer(df,col),statistic = 'D',permutation.test = T,B=5000)

  pvalues$summer<-data.frame(Statistic=c('U','L','D'),
                              p.values=c(U$p.value,L$p.value,D$p.value))

  U<-foster.test(autumn(df,col),statistic = 'U',permutation.test = T,B=5000)
  L<-foster.test(autumn(df,col),statistic = 'L',permutation.test = T,B=5000)
  D<-foster.test(autumn(df,col),statistic = 'D',permutation.test = T,B=5000)

  pvalues$autumn<-data.frame(Statistic=c('U','L','D'),
                              p.values=c(U$p.value,L$p.value,D$p.value))

  return(pvalues)
}

estMadTx<-pval(tmax,3) #tiempo compilación alto
estMadTm<-pval(tmin,3)

#graficos (para tortosa cambiar fechas de inicio y modificar funciones estaciones
#para concordar con el dataframe)
graf<-function(data,nombre,tipo){
  title<-paste(nombre,'-',tipo)
  breaks<-seq(1, nrow(data), by = 20)
  labels<-seq(1920, 2020, by = 20)

```

```

S1<-ggplot2::scale_x_continuous(breaks = breaks,
                               labels = labels,
                               sec.axis = ggplot2::sec_axis(~ 2021 - .,
                                                            name = "Year (backward)",
                                                            breaks = 1919 + breaks))

S2<-ggplot2::scale_x_continuous(breaks = breaks,
                               labels = labels)
COL<-ggplot2::theme(axis.title.x = ggplot2::element_text(colour = "red"),
                   axis.text.x = ggplot2::element_text(colour = "red"),
                   axis.title.x.top = ggplot2::element_text(colour = "blue"),
                   axis.text.x.top = ggplot2::element_text(colour = "blue"))

w<-function(t) ifelse(t==1,0,t^2/(t-1))

#Nplot
g1<-N.plot(data,weights=w)
g1$labels$title<-title
g1$labels$subtitle<-NULL
g1$labels$x<- 'Year (forward)'
g1<-g1+S1+COL

print(g1)
}

graf_final<-function(df,col,tipo){

  g1<-graf(winter(df,col),'Invierno',tipo)
  g2<-graf(spring(df,col),'Primavera',tipo)
  g3<-graf(summer(df,col),'Verano',tipo)
  g4<-graf(autumn(df,col),'Otoño',tipo)

  g<-ggpubr::ggarrange(g1,g2,g3,g4,
                      nrow = 2,ncol=2, common.legend = TRUE, legend = "bottom")

  print(g)
}

graf_final(tmax,3,'Tmax')

```

Mapas

```

library(sf)
library(ggplot2)
library(rnaturalearth)
library(rnaturalearthdata)
library(dplyr)
library(ggpubr)
library(rgeos)

####MAPAS CIUDADES

```



```

spain <- ne_countries(scale = "medium", returnclass = "sf") %>%
  filter(admin == "Spain")
cities <- data.frame(
  city = c("Castellón", "Madrid", "Daroca", "Cáceres", "Huelva", "Tortosa"),
  Lon = c(-0.04935, -3.684500, -1.4143600, -6.37224, -6.94004, 0.5216000),
  Lat = c(39.98567, 40.415260, 41.1145800, 39.47649, 37.26638, 40.8124900)
)

cities_sf <- st_as_sf(cities, coords = c("Lon", "Lat"), crs = 4326)

city_colors <- c("#E41A1C", "blue", "#4DAF4A", "black", "#FF7F00", "purple")
names(city_colors) <- cities$city

# Plot del mapa de España y ciudades
p <- ggplot() +
  geom_sf(data = spain) +
  geom_sf(data = cities_sf, aes(color = city), size = 3.5) +
  scale_color_manual(values = city_colors) +
  theme_minimal() +
  labs(title = "Ciudades en España", color = "Ciudad")

p <- p +
  coord_sf(xlim = c(-10, 5), ylim = c(35, 45))

print(p)

###MAPAS CON P-VALORES

#obtencion p-valores para cada ciudad y almacenar vector
pvalD<-function(df,col,stat){
  uncor<-split_uncor(df,col)

  uncor2<-series_untie2(uncor)

  w<-function(t) ifelse(t==1,0,t^2/(t-1))

  D<-foster.test(uncor2,weights=w,
    statistic = stat)

  D<-D$p.value
  return(D)
}

pval<-function(df1,df2,stat){
  pvalues<-c(pvalD(df1,2,stat=stat),
    pvalD(df1,3,stat=stat),
    pvalD(df1,4,stat=stat),
    pvalD(df1,6,stat=stat),
    pvalD(df1,7,stat=stat),
    pvalD(df2,2,stat=stat))

  return(pvalues)
}

```

```

pvDTx<-pval(tmax,tortosa_tmax,'D')
pvDTm<-pval(tmin,tortosa_tmin,'D')

mapSpain <- function(Z, coords, ref, leyenda, nombre,niveles) {
  spain <- ne_countries(scale = "medium", returnclass = "sf") %>%
    filter(admin == "Spain")

  data <- cbind(coords, Z)

  data_sf <- st_as_sf(data, coords = c("Longitude", "Latitude"), crs = 4326)

  # mapA
  colores<-c("#CB181D", "#F8968C", "#EFB6B0", "#86CAE1", "#81B1D6", "#5F6DAE", "#023FA5")
  color_map<-setNames(colores,niveles)
  p <- ggplot(data = spain) +
    geom_sf() +
    geom_sf(data = data_sf, aes(color = Z), size = 3.5) +
    theme_minimal() +
    scale_color_manual(values=color_map)+
    labs(title = nombre, color = "P-Value")

  p <- p +
    coord_sf(xlim = c(-10, 5), ylim = c(35, 45))

  # LEYENDA
  if (leyenda) {
    p <- p + theme(legend.position = "bottom")
  } else {
    p <- p + theme(legend.position = "none") # Ocultar leyenda si leyenda es FALSE
  }

  print(p)
}

# cortes y niveles
cortes <- c(0, 10^(-10), 10^(-5), 10^(-2), 0.05,0.1,0.2 ,1)
niveles <- c("[0, 10e-10)", "[10e-10, 10e-05)", "[10e-05, 0.01)", "[0.01, 0.05)",
  "[0.05, 0.1)", "[0.1, 0.2)", "[0.2, 1)")

colores<-c("#CB181D", "#F8968C", "#EFB6B0", "#86CAE1", "#81B1D6", "#5F6DAE", "#023FA5")

color_map<-setNames(colores,niveles)

# coordenadas ciudades
Lat<-c(39.98567,40.415260,41.1145800, 39.47649,37.26638 , 40.8124900)
Lon<-c(-0.04935,-3.684500,-1.4143600, -6.37224, -6.94004, 0.5216000 )
# Cortar los p-valores en los intervalos definidos
ppDTx <- cut(pvDTx, breaks = cortes, include.lowest = TRUE, labels = niveles)

ppDTm <- cut(pvDTm, breaks = cortes, include.lowest = TRUE, labels = niveles)

map1<-mapSpain(Z = ppDTx, coords = data.frame(Longitude = Lon, Latitude = Lat),
  ref = 0.05,leyenda = T,nombre='Temperaturas Máximas',niveles=niveles)

```

```

map2<-mapSpain(Z = ppDTm, coords = data.frame(Longitude = Lon, Latitude = Lat),
               ref = 0.05,leyenda = T,nombre='Temperaturas Mínimas',niveles=niveles)

#mostrar juntos
g<-ggpubr::ggarrange(map1,map2,
                     nrow = 1,ncol=2, common.legend = F, legend = 'bottom')

print(g)

```


Bibliografía

- [1] CHANDLER, K.N. (1952). *The Distribution and Frequency of Record Values*. Journal of the Royal Statistical Society. Series B (Methodological), 14 (2), 220-228. doi: 10.2307/2983870
- [2] ARNOLD, B. C., BALAKRISHNAN, N., & NAGARAJA, H. N. (2008). *A First Course in Order Statistics*. Society for Industrial and Applied Mathematics.
- [3] ARNOLD, B. C., BALAKRISHNAN, N., & NAGARAJA, H. N. (1998). *Records*. New York: Wiley.
- [4] RÉNYI, A. (1962). Théorie des éléments saillants d'une suite d'observations. *Annales scientifiques de l'Université de Clermont-Ferrand 2. Série Mathématiques*, 8 (2), 7-13.
- [5] NEVZOROV, V.B. (2001). *Records: Mathematical Theory*. Rhode Island: American Mathematical Society.
- [6] WERGEN, G. (2013). Records in stochastic processes - Theory and Applications. *Journal of Physics A: Mathematical and Theoretical*, 46 (22), 223001. doi: 10.1088/1751-8113/46/22/223001
- [7] CEBRIÁN, A.C., CASTILLO-MATEO, J., ASÍN J. (2022). *Record tests to detect non-stationarity in the tails with an application to climate change*. Stochastic Environmental Research and Risk Assesment, 36 (2), 313-330. doi: 10.1007/s00477-021-02122-w
- [8] CASTILLO-MATEO, J., CEBRIÁN, A.C., ASÍN J. (2023). *RecordTest: An R package to analyse non-stationarity in the extremes based on record-breaking events*. Journal of Statistical Software, 106 (5). doi: 10.18637/jss.v106.i05
- [9] CASTILLO-MATEO, J., CEBRIÁN, A.C., ASÍN J. (2023). *Statistical analysis of extreme and record-breaking daily maximum temperatures in peninsular Spain during 1960–2021*. Atmospheric Research, 293. doi: 10.1016/j.atmosres.2023.106934
- [10] DIERSEN, J., TRENKLER, G. (1996). *Records tests for trend in location*. Statistics, 28 (1), 1-12. doi: 10.1080/02331889708802543
- [11] DIERSEN, J., TRENKLER, G. (2001). *Weighted record tests for splitted series of observations*. In: Kunert J., Trenkler G. (eds) Mathematical statistics with applications in biometry: festschrift in honour of Prof. Dr. Siegfried Schach, Lohmar: Josef Eul Verlag, pp 163-178.
- [12] FOSTER, F.G., STUART, A. (1954). *Distribution-free tests in time-series based on the breaking of records*. Journal of the Royal Statistical Society: Series B (Methodological), 16 (1), 1-13. doi: 10.1111/j.2517-6161.1954.tb00143.x
- [13] KING, M., WU, P. (1997). *Locally optimal one-sided tests for multiparameter hypotheses*. Econometrics Reviews, 16 (2), 131-156. doi: 10.1080/07474939708800379
- [14] WELCH, W.J. (1990). *Construction of permutation tests*. Journal of the American Statistical Association, 85 (411), 693-698. doi: 10.1080/01621459.1990.10474929

- [15] JOHNSON, N.L., KEMP, A.W., KOTZ, S. (2005). Preliminary Information. En D.J. Balding et al. (Eds). *Univariate Discrete Distributions*, 1-73. Hoboken, New Jersey: Wiley.
- [16] CUZZOCREA, A., FADDA, E., BALDO, A. (2021) *Lyapunov Central Limit Theorem: Theoretical Properties and Applications in Big-Data-Populated Smart City Settings*. ACM International Conference Proceeding Series, 34-38. doi: 10.1145/3481646.3481652
- [17] BILLINGSLEY, P. (1995) *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics, 3rd ed, pp 359-362.
- [18] *El clima en España*. https://educativo.ign.es/atlas-didactico/clima-eso/el_clima_en_espaa.html
- [19] *Clasificación climática de Koppen*. <http://meteo.navarra.es/definiciones/koppen.cfm>