



FACULTAD DE ESTUDIOS ESTADÍSTICOS

MÁSTER EN MINERÍA DE DATOS E INTELIGENCIA DE NEGOCIOS

Curso 2017/2018

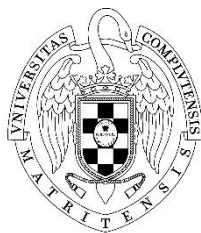
Trabajo de Fin de Máster

**TÍTULO: ANÁLISIS DE LAS VARIACIONES DEL
CLIMA EN LA ESPAÑA PENINSULAR**

Alumna: Belén Caro Poza

Tutor: José Luis Valencia Delfa

Septiembre de 2018



UNIVERSIDAD COMPLUTENSE
MADRID

ÍNDICE

1.	INTRODUCCIÓN	6
2.	OBJETIVOS	7
3.	FUENTES DE DATOS.....	8
3.1.	Naturaleza de los datos.....	8
3.2.	Reconstrucción de las series	9
4.	METODOLOGÍA EMPLEADA.....	11
4.1.	Procedimientos de homogeneización.....	11
4.2.	Análisis de series temporales	13
4.2.1.	Procesos estocásticos estacionarios	13
4.2.2.	Modelos autorregresivos AR(p)	14
4.2.3.	Modelos de medias móviles MA(q).....	14
4.2.4.	Modelos mixtos ARMA(p,q)	14
4.2.5.	Procesos integrados: Modelos ARIMA(p,d,q)	15
4.2.6.	Modelos estacionales ARIMA(p,d,q)(P,D,Q)s.....	15
4.2.7.	Metodología Box-Jenkins	15
4.3.	Análisis de tendencias	17
4.3.1.	Técnica del preblanqueo	17
4.3.2.	Contrastes de Mann-Kendall y Theil-Sen	18
4.3.3.	Regresión lineal	18
5.	DESARROLLO DEL TRABAJO Y RESULTADOS	20
5.1.	Homogeneización de las series	20
5.1.1.	Preparación de los ficheros de entrada	20
5.1.2.	Primer análisis exploratorio de los datos diarios originales para el control de calidad	21
5.1.3.	Agregación de los datos diarios en series mensuales	24
5.1.4.	Homogeneización de las series mensuales	24
5.1.5.	Homogeneización de las series diarias con los puntos de corte mensuales... ..	26
5.2.	Análisis de series temporales	27
5.2.1.	Representación gráfica y descomposición estacional.....	28
5.2.2.	Análisis de la estacionariedad de las series y ajuste de los modelos ARIMA..	30
5.2.3.	Cálculo y representación de las predicciones	43
5.3.	Análisis de la tendencia.....	45
5.3.1.	Cálculo y representación de la tendencia en mapas.....	45
5.3.2.	Regresión lineal	50

6.	CONCLUSIONES	53
7.	BIBLIOGRAFÍA.....	55
8.	ANEXOS	57
A.	Tablas	57
B.	Códigos.....	61

ÍNDICE DE FIGURAS

Figura 1. Evolución del número de estaciones meteorológicas (1920-2012).....	8
Figura 2. Diagrama de flujo del funcionamiento de Climatol, mostrando sus procesos iterativos	12
Figura 3. Correlaciones entre las series (izquierda) y mapa de las estaciones (derecha; los colores identifican grupos de estaciones con variabilidad similar)	22
Figura 4. Anomalías de una serie homogénea (izquierda) y una heterogénea (derecha)	23
Figura 5. Histogramas de anomalías estandarizadas	23
Figura 6. Gráfico de la calidad/singularidad de las series finales.....	24
Figura 7. Histogramas de valores máximos de SNHT encontrados en ventanas escalonadas superpuestas (izquierda) y en las series completas (derecha)	25
Figura 8. Histogramas de valores máximos de SNHT encontrados en las series completas	25
Figura 9. Inhomogeneidades de una serie detectadas en la etapa 1.....	26
Figura 10. Reconstrucción de la serie a partir de los tres subconjuntos homogéneos	26
Figura 11. Representación gráfica y descomposición estacional de la temperatura máxima....	28
Figura 12. Representación gráfica y descomposición estacional de la temperatura mínima	29
Figura 13. Representación gráfica y descomposición estacional de la temperatura media.....	29
Figura 14. Representación gráfica y descomposición estacional de las precipitaciones	30
Figura 15. Funciones de autocorrelación simple y parcial de la temperatura máxima	31
Figura 16. Funciones de autocorrelación simple y parcial de la temperatura máxima con una diferencia en la parte regular	31
Figura 17. Funciones de autocorrelación simple y parcial de la temperatura máxima con una diferencia en la parte estacional	32
Figura 18. Funciones de autocorrelación simple y parcial de la temperatura máxima con un autorregresivo y una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional.....	33
Figura 19. Funciones de autocorrelación simple y parcial de la temperatura máxima con una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional	34
Figura 20. Diagramas de normalidad de los residuos de la temperatura máxima	35
Figura 21. Funciones de autocorrelación simple y parcial de la temperatura mínima con una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional	36
Figura 22. Diagramas de normalidad de los residuos de la temperatura mínima	36
Figura 23. Funciones de autocorrelación simple y parcial de la temperatura media con una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional	37
Figura 24. Diagramas de normalidad de los residuos de la temperatura media	37
Figura 25. Representación gráfica de los correlogramas simple y parcial de las precipitaciones	38
Figura 26. Funciones de autocorrelación simple y parcial de las precipitaciones con una diferencia en la parte estacional.....	39
Figura 27. Funciones de autocorrelación simple y parcial de las precipitaciones con un autorregresivo y una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional.....	40
Figura 28. Funciones de autocorrelación simple y parcial de las precipitaciones con una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional	41
Figura 29. Diagramas de normalidad de los residuos de las precipitaciones	42

Figura 30. Residuales y predicciones de la temperatura máxima.....	43
Figura 31. Residuales y predicciones de la temperatura mínima	43
Figura 32. Residuales y predicciones de la temperatura media	44
Figura 33. Residuales y predicciones de las precipitaciones totales.....	44
Figura 34. Gráfico de dispersión entre la temperatura máxima estimada y la real (izquierda) y entre la temperatura mínima estimada y la real (derecha).....	45
Figura 35. Gráfico de dispersión entre la temperatura media estimada y la real (izquierda) y entre las precipitaciones estimadas y las reales (derecha).....	45
Figura 36. Tendencia de la temperatura máxima en la España peninsular en el mes de enero desde 1950 hasta 2011	47
Figura 37. Tendencia de la temperatura máxima en la España peninsular en cada uno de los meses del año (ordenados de enero a diciembre) desde 1950 hasta 2011	48
Figura 38. Tendencia de la temperatura mínima en la España peninsular en cada uno de los meses del año (ordenados de enero a diciembre) desde 1950 hasta 2011	49
Figura 39. Tendencia de la temperatura media en la España peninsular en cada uno de los meses del año (ordenados de enero a diciembre) desde 1950 hasta 2011	50

ÍNDICE DE TABLAS

Tabla 1. Periodo en el que está operativa cada estación meteorológica (1920-2012).....	10
Tabla 2. Estaciones meteorológicas operativas en ambos periodos	28
Tabla 3. Estimaciones y significatividad de los parámetros del modelo ARIMA ajustado	33
Tabla 4. Correlaciones entre los parámetros del modelo ARIMA ajustado	34
Tabla 5. Estimaciones y significatividad de los parámetros del modelo ARIMA ajustado	34
Tabla 6. Contrastes de aleatoriedad de los residuos del modelo ARIMA ajustado	35
Tabla 7. Parámetros estimados de los modelos ajustados para la temperatura máxima, mínima y media, respectivamente.....	38
Tabla 8. Estimaciones y significatividad de los parámetros del modelo ARIMA ajustado	40
Tabla 9. Correlaciones entre los parámetros del modelo ARIMA ajustado	41
Tabla 10. Estimaciones y significatividad de los parámetros del modelo ARIMA ajustado	41
Tabla 11. Contrastes de aleatoriedad de los residuos del modelo ARIMA ajustado	42
Tabla 12. Parámetros estimados del modelo ajustado para las precipitaciones.....	42
Tabla 13. Estimadores de los parámetros de la regresión lineal	51
Tabla 14. Estaciones meteorológicas del periodo 1950-2011	58
Tabla 15. Inhomogeneidades detectadas en las series del periodo 1950-2011 para la temperatura máxima y mínima.....	59
Tabla 16. Inhomogeneidades detectadas en las series del periodo 1950-2011 para la temperatura media y precipitaciones.....	60
Tabla 17. Inhomogeneidades detectadas en las series del periodo 2014-2017 para la temperatura máxima y media.....	60
Tabla 18. Inhomogeneidades detectadas en las series del periodo 2014-2017 para la temperatura mínima y precipitaciones.....	60

1. INTRODUCCIÓN

La Organización Meteorológica Mundial (OMM) define el clima como la media de las condiciones meteorológicas en una zona determinada durante un largo período de tiempo. Se ha establecido por convenio que el período en que se calcula la media de las condiciones meteorológicas es de 30 años.

El denominado cambio climático obedece a las variaciones climáticas que han tenido lugar durante el último siglo, y la velocidad a la que se están produciendo es un hecho excepcional. En la actualidad existe unanimidad entre los expertos del cambio climático al determinar que se está produciendo un calentamiento general del clima sin precedentes. Las olas de calor, las inundaciones, las sequías y las tormentas serán cada vez más frecuentes e intensas como resultado de este cambio climático.

El estudio del clima se basa en los datos meteorológicos. Para ello, la Agencia Estatal de Meteorología (AEMET) dispone de una red de estaciones meteorológicas que efectúan sistemáticamente observaciones para tomar datos y caracterizar los elementos del clima. Las mediciones climáticas registradas por las estaciones pueden verse afectadas por eventos ajenos a los factores climáticos y al paso del tiempo, como pueden ser cambios en la instrumentación, errores en la toma de medidas o en su transmisión, reubicaciones o cambios en el medio ambiente. Estos cambios producidos en el entorno pueden deberse, por ejemplo, al crecimiento urbano, a incendios o a cambios en el uso del suelo. Estas alteraciones en las mediciones de los diferentes elementos del clima producen inhomogeneidades en las series, lo cual supone que las conclusiones obtenidas sean erróneas. Para abordar este problema, se ha desarrollado un proceso, llamado homogeneización, que trata de eliminar estas perturbaciones no deseadas y permite que las series climatológicas revelen solo las variaciones climáticas.

España posee una gran variedad climática, además de una elevada variabilidad interanual y una notable amplitud de valores diarios extremos. Los análisis realizados sobre las tendencias de temperatura confirman que se ha producido un incremento de la temperatura media anual desde mediados de los años 70, siendo el calentamiento más importante en verano.

En el presente trabajo se estudiarán las tendencias climáticas ocurridas en el territorio español peninsular en las últimas décadas (1950-2011).

2. OBJETIVOS

El objetivo general de este estudio consiste en lograr anticiparse a las condiciones meteorológicas de la España peninsular, analizando su comportamiento climático desde el año 1950.

Para ello, se plantean varios objetivos específicos:

- Generar un conjunto de datos homogeneizado y depurado para el periodo 1950-2011 y otro para el periodo 2014-2017.
- Observar la evolución en el tiempo de la temperatura máxima, mínima y media y de las precipitaciones con el fin de ajustar un modelo de series temporales automático para cada una de ellas, que sea válido para todas las estaciones meteorológicas y que permita conocer su comportamiento futuro.
- Calcular y analizar la tendencia que han seguido las precipitaciones y la temperatura máxima, mínima y media en cada mes del año a nivel provincial, para conocer con mayor precisión su evolución a lo largo de los años.
- Examinar la influencia que tiene la localización geográfica de las estaciones meteorológicas sobre las temperaturas y las precipitaciones.

En base a estos objetivos, surgen las siguientes hipótesis:

- Es posible predecir la temperatura mínima, máxima y media y las precipitaciones de un lugar de la España peninsular en un determinado periodo de tiempo.
- La temperatura en España tiene una tendencia creciente.
- Las precipitaciones han ido disminuyendo a lo largo de los años en España.
- La localización de las estaciones meteorológicas influye sobre la tendencia que han seguido las temperaturas y las precipitaciones en la España peninsular en los últimos años.

3. FUENTES DE DATOS

3.1. Naturaleza de los datos

A pesar de que en España existe un gran número de estaciones meteorológicas, lamentablemente, la mayoría de ellas han empezado a operar a mediados del año 2013. La Agencia Estatal de Meteorología (AEMET) proporciona datos históricos diarios desde enero de 1920 hasta febrero de 2012 de un total de 114 estaciones meteorológicas repartidas por toda España, pero, desde mediados del año 2013, el número de estaciones operativas asciende a 827.

Por tanto, para realizar el presente estudio se dispone de datos desde el 1 de enero de 1920 hasta el 29 de febrero de 2012 y desde el 7 de mayo de 2013 hasta el 31 de mayo de 2017. Los datos pertenecientes al periodo intermedio no son abiertos, ya que son de pago.

Durante ese primer periodo comprendido entre 1920 y 2012, el número de estaciones operativas en las primeras décadas era muy bajo. Este número fue aumentando con el paso de los años, llegando a alcanzar un máximo de 94 estaciones con datos disponibles en el año 1985 (figura 1).

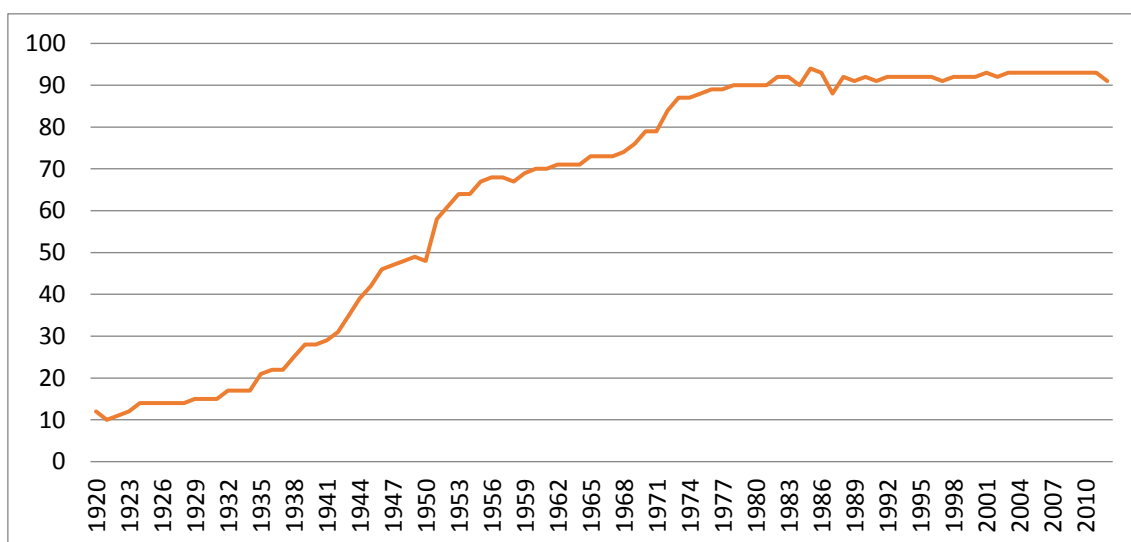


Figura 1. Evolución del número de estaciones meteorológicas (1920-2012)

En cambio, desde mediados del año 2013, prácticamente todas las estaciones tienen datos durante todo el periodo.

Dado que este estudio está enfocado al clima de la España peninsular, los datos pertenecientes a las estaciones de Islas Baleares, Islas Canarias, Ceuta y Melilla serán descartados, de modo que el número de observatorios utilizados en el primer periodo serán 98 y, en el segundo periodo, 716.

Las variables relacionadas con la localización geográfica de las estaciones son el municipio y la provincia, la altitud medida en metros y la latitud y longitud medidas en coordenadas sexagesimales.

Las variables relacionadas con los elementos del clima disponibles diariamente para cada estación son las siguientes:

- Datos de temperatura del aire: Temperatura máxima (°C), temperatura mínima (°C) y temperatura media (°C).
- Precipitaciones (l/m²).
- Horas de sol.
- Datos de viento: Racha de viento máxima (km/h) y velocidad media del viento (km/h).
- Datos de presión atmosférica: Presión máxima (mbar) y presión mínima (mbar).

Este estudio se centrará en el análisis de las temperaturas máximas, mínimas y medias y las precipitaciones.

3.2. Reconstrucción de las series

Durante el periodo 1920-2012, los datos climáticos recogidos por los diferentes observatorios son muy irregulares, ya que abarcan periodos de tiempo diferentes, lo que dificulta su análisis y tratamiento. Sin embargo, en algunos casos, el cese de una estación está acompañado del comienzo de otra en un lugar lo suficientemente cercano como para formar un solo conjunto de series más largo a partir de ellas. Si el emplazamiento de estos observatorios es próximo, normalmente, las diferencias en las temperaturas y precipitaciones deberán ser pequeñas. En estos casos, se va a construir una nueva serie a partir de las demás con el fin de obtener un conjunto de datos más completo, la cual será atribuida a la última estación en activo.

Se han reconstruido un total de 17 series de diferentes formas. En cuanto a las estaciones con identificador 0367, 1484C, 1690A, 2444, 2465, 3168C, 4121, 4642E, 5270B, 8500A, 9381I y 9771C, la reconstrucción se ha realizado simplemente uniendo cada una de ellas con su estación complementaria, ya que sus periodos no coincidían en ningún momento. La estación con código 1208H se ha reconstruido a partir de la unión de la misma con sus dos estaciones complementarias. Por último, en cuanto a las estaciones 3260B, 3469A, 6325O y 9091O, una parte del periodo en el que están operativas coincide con el de sus estaciones complementarias, por lo que la reconstrucción de esos periodos se ha realizado promediando los datos de ambos observatorios.

En la tabla 1 se detallan las fechas de inicio y fin del periodo en el que está operativa cada una de las estaciones, incluidas las series reconstruidas, así como la provincia a la que pertenecen. La gran mayoría finalizan en el año 2012, pero solo unas pocas tienen datos desde el comienzo del periodo.

ID	PROVINCIA	INICIO	FIN	ID	PROVINCIA	INICIO	FIN
0016A	Tarragona	01/02/45	29/02/12	3200	Madrid	01/01/51	29/02/12
0076	Barcelona	01/03/24	29/02/12	3260B	Toledo	01/01/20	29/02/12
0200E	Barcelona	01/01/20	31/01/12	3469A	Cáceres	01/01/20	29/02/12
0367	Girona	01/01/20	29/02/12	4121	Ciudad Real	01/01/20	29/02/12
1014	Gipuzkoa	18/08/55	29/02/12	4452	Badajoz	01/01/55	29/02/12
1024E	Gipuzkoa	01/01/28	29/02/12	4642E	Huelva	01/01/20	29/02/12
1082	Bizkaia	13/03/47	29/02/12	5270B	Jaén	05/01/20	29/02/12
1109	Cantabria	12/09/53	29/02/12	5402	Córdoba	04/01/59	29/02/12
1110	Cantabria	01/01/24	07/04/97	5514	Granada	12/01/37	29/02/12
1111	Cantabria	01/06/46	29/02/12	5530E	Granada	07/01/72	29/02/12
1208H	Asturias	01/10/38	29/02/12	5783	Sevilla	01/01/51	29/02/12
1212E	Asturias	01/08/68	29/02/12	5796	Sevilla	02/01/46	29/02/12
1249I	Asturias	01/12/72	29/02/12	5910	Cádiz	09/01/88	29/02/12
1387	A Coruña	01/10/30	29/02/12	5960	Cádiz	02/01/46	29/02/12
1387E	A Coruña	01/12/71	29/02/12	5973	Cádiz	11/01/35	29/02/12
1428	A Coruña	01/11/43	29/02/12	6001	Cádiz	05/01/45	29/02/12
1484C	Pontevedra	01/10/63	29/02/12	6155A	Málaga	05/01/42	29/02/12
1495	Pontevedra	01/05/56	29/02/12	6325O	Almería	01/01/34	29/02/12
1505	Lugo	01/01/51	29/02/12	7031	Murcia	08/01/44	29/02/12
1549	León	01/01/51	29/02/12	7178I	Murcia	04/01/84	29/02/12
1690A	Ourense	01/01/52	29/02/12	7228	Murcia	01/01/40	29/02/12
2030	Soria	01/11/43	29/02/12	8019	Alicante	01/01/67	29/02/12
2331	Burgos	01/08/43	29/02/12	8025	Alicante	09/01/38	29/02/12
2401	Palencia	01/01/51	31/07/90	8096	Cuenca	01/01/51	29/02/12
2422	Valladolid	01/10/73	29/02/12	8175	Albacete	12/01/39	29/02/12
2444	Ávila	01/01/53	29/02/12	8178D	Albacete	01/01/83	29/02/12
2462	Madrid	01/01/46	29/02/12	8368U	Teruel	04/01/86	29/02/12
2465	Segovia	01/01/31	29/02/12	8414A	Valencia	01/01/65	29/02/12
2539	Valladolid	01/11/36	29/02/12	8416	Valencia	11/01/37	29/02/12
2614	Zamora	01/02/20	29/02/12	8500A	Castellón	01/01/37	29/02/12
2661	León	01/01/38	29/02/12	9091O	Álava	03/01/43	29/02/12
2867	Salamanca	01/01/45	29/02/12	9170	La Rioja	11/01/48	29/02/12
2870	Salamanca	01/01/70	29/02/12	9262	Navarra	09/01/53	29/02/12
3013	Guadalajara	01/09/49	29/02/12	9263D	Navarra	01/01/75	29/02/12
3129	Madrid	01/01/51	29/02/12	9381I	Teruel	01/01/42	31/12/11
3168C	Guadalajara	01/01/51	29/02/12	9390	Zaragoza	01/01/20	29/02/12
3175	Madrid	01/01/61	29/02/12	9434	Zaragoza	01/01/41	29/02/12
3191E	Madrid	01/01/78	29/02/12	9771C	Lleida	06/01/59	29/02/12
3195	Madrid	01/01/20	29/02/12	9898	Huesca	03/01/43	29/02/12
3196	Madrid	01/05/45	29/02/12	9981A	Tarragona	03/01/43	29/02/12

Tabla 1. Periodo en el que está operativa cada estación meteorológica (1920-2012)

4. METODOLOGÍA EMPLEADA

4.1. Procedimientos de homogeneización

A lo largo de los años se han desarrollado diferentes procedimientos de homogeneización que permiten eliminar o reducir en la medida de lo posible las alteraciones producidas en los datos climáticos a causa de factores externos.

El procedimiento utilizado en el presente estudio consiste en comparar una serie problema con otra de referencia creada a partir de la estimación de otras series seleccionadas por su proximidad o elevada correlación, con lo que se eliminarían las posibles inhomogeneidades de la serie problema. Una vez construida la serie de referencia, la comparación se realiza a partir de la serie de diferencias entre ambas, la cual carecerá de la estacionalidad propia de las series climáticas. Por tanto, esta serie debería comportarse como un ruido blanco, de tal modo que se pueda considerar como inhomogeneidad a cualquier anomalía que se aleje del mismo.

El paquete de R *Climatol* va a ser empleado para llevar a cabo este proceso, ya que contiene las funciones necesarias para la correcta homogeneización de las series climáticas. Además, permite controlar la calidad de los datos, así como rellenar los valores ausentes.

A la hora de completar estos datos faltantes, *Climatol* permite seleccionar el tipo de normalización más adecuado para cada tipo de serie. Las opciones disponibles son las siguientes, siendo m_X la media y s_X la desviación típica de una serie X :

1. Restar la media: $x = X - m_X$
2. Dividir por la media: $x = X/m_X$
3. Estandarizar: $x = (X - m_X)/s_X$

Cuando se trabaja con variables limitadas por cero y con una distribución de probabilidad sesgada, como la precipitación o la velocidad del viento, la normalización por proporción respecto a la media (caso 2) es preferible a la estandarización completa (caso 3), la cual debe aplicarse, por ejemplo, a las series de temperatura.

El principal problema de esta metodología es que las medias y desviaciones típicas de las series no se conocen si estas no están completas, que es lo más habitual, por lo que *Climatol* las calcula, en primer lugar, con los datos disponibles en cada serie y rellena los valores ausentes a partir de estas medias y desviaciones típicas provisionales, volviendo a calcularlas con las series completas. Posteriormente, se vuelven a calcular los datos que en un principio estaban ausentes con los nuevos parámetros, lo que dará lugar a nuevas medias y desviaciones típicas. Este proceso se repite sucesivamente hasta que ninguna de las medias cambie.

Una vez estabilizadas las medias, se normalizan todos los datos y se procede a estimarlos de la siguiente forma, siendo \hat{y} el valor estimado en función de los n datos x_j más próximos disponibles en cada momento del tiempo y sus correspondientes pesos w_j :

$$\hat{y} = \frac{\sum_{j=1}^{j=n} w_j x_j}{\sum_{j=1}^{j=n} w_j}$$

Estos datos estimados constituyen las series de referencia que han de ser comparadas con sus correspondientes series observadas. Para ello, se calcula la serie de diferencias entre ambas, la cual se corresponde con la serie de anomalías, que permitirá controlar la calidad de los datos y eliminar las anomalías que superen un umbral predefinido, así como comprobar su homogeneidad mediante la aplicación del Test de Homogeneidad Normal Estándar (SNHT, por sus siglas en inglés).

Cuando los valores máximos SNHT de las series superan un determinado umbral, la serie se divide por el punto de máximo SNHT, pasando todos los datos antes del cambio a una nueva serie con las mismas coordenadas, pero añadiendo un sufijo numérico al código y al nombre de la estación. Este procedimiento se realiza sucesivamente, partiendo únicamente las series con mayores valores SNHT en cada ciclo, hasta que no se encuentren más inhomogeneidades.

El problema de esta prueba radica en que tan solo busca un punto de ruptura en cada serie, por lo que la existencia de dos o más saltos en la media de un tamaño similar podría no ser detectado. Para solucionarlo, en primer lugar, se aplica el test sobre ventanas temporales solapadas y, posteriormente, se aplica a las series completas, momento en el que la prueba tiene más poder de detección. Finalmente, se rellenan todos los datos ausentes de todas las series y subseries homogéneas con el mismo procedimiento de estimación explicado anteriormente.

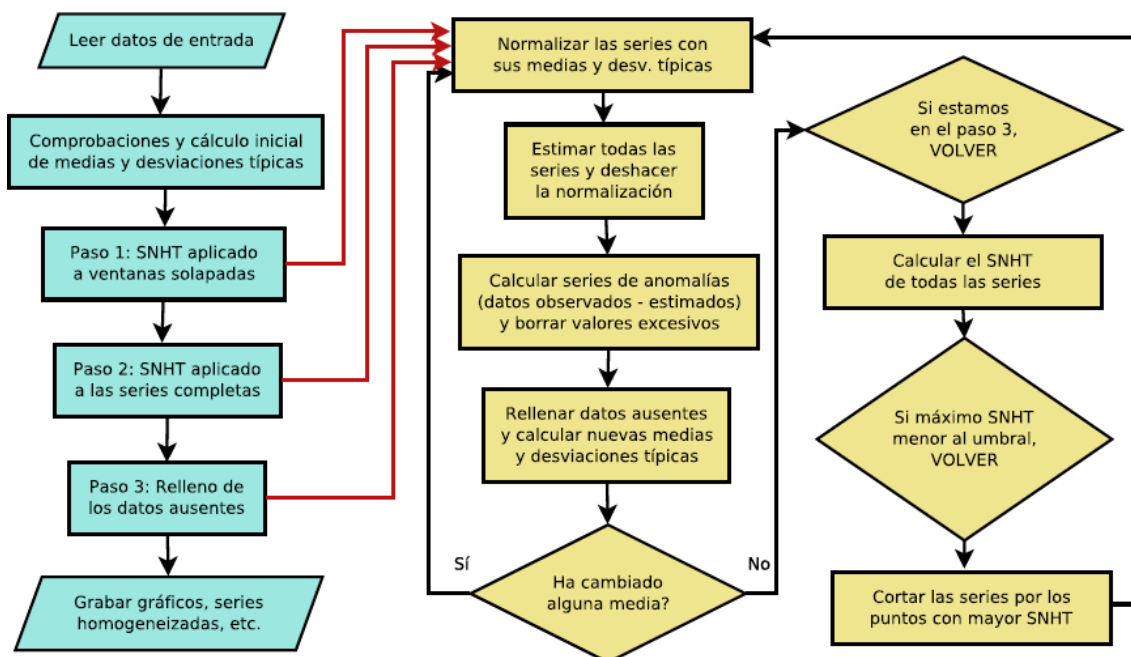


Figura 2. Diagrama de flujo del funcionamiento de Climatol, mostrando sus procesos iterativos

4.2. Análisis de series temporales

El análisis de una serie temporal es el resultado de observar la distribución de los valores de una variable a lo largo del tiempo en intervalos regulares (diarios, mensuales, anuales...). Un modelo de series temporales debe reproducir las características de esa variable a partir de su comportamiento pasado para predecir sus valores futuros. Existen diferentes técnicas de modelado de series, como aquellas basadas en métodos de suavizado o en la metodología Box-Jenkins. Estas últimas serán las empleadas en este estudio y, para ello, se utilizará SAS Base.

4.2.1. Procesos estocásticos estacionarios

Un proceso estocástico x_t es el conjunto de variables aleatorias relacionadas entre sí ordenadas en función del tiempo. En el análisis de series temporales se debe determinar qué procesos estocásticos permiten generar las series con el fin de caracterizar sus comportamientos pasados y predecir sus valores futuros. Estos procesos deben tener una cierta estabilidad en el tiempo, ya que, en caso contrario, no serán capaces de predecir las series. A esta condición de estabilidad se le conoce como estacionariedad.

Para desarrollar la metodología Box-Jenkins, un proceso estocástico debe ser estacionario en sentido amplio o débil:

- Media constante: $\mu_t = \mu \quad \forall t$
- Varianza constante: $\sigma_t^2 = \sigma^2 \quad \forall t$
- La covarianza entre dos variables de la serie solo depende de su separación en el tiempo: $cov(x_t, x_{t+k}) = E[(x_t - \mu)(x_{t-k} - \mu)] = \gamma_k \quad \forall k$

El proceso estocástico estacionario más sencillo es el proceso de ruido blanco, cuya media es cero, su varianza es constante y sus variables están incorreladas:

$$E[x_t] = 0, \quad var[x_t] = \sigma^2 \quad y \quad \gamma_k = 0 \quad \forall k = 1, 2, \dots$$

El coeficiente de autocorrelación permite identificar si una serie es estacionaria, ya que mide el grado de dependencia lineal entre las variables de la serie. Se define de la siguiente forma:

$$\rho_k = \frac{cov(x_t, x_{t-k})}{\sqrt{var(x_t)var(x_{t-k})}} = \frac{\gamma_k}{\gamma_0} \quad \gamma_0 = \sigma_{x_t}^2 \quad \forall t$$

La función de autocorrelación simple (ACF, por sus siglas en inglés) es el conjunto de coeficientes de autocorrelación expresados en función del retardo. El correlograma de esta función, unido al de la función de autocorrelación parcial (PACF, por sus siglas en inglés), va a permitir identificar el tipo de modelo a ajustar en el caso de que la serie no sea estacionaria. Estos modelos pueden ser autorregresivos (AR(p)), medias móviles (MA(q)) o mixtos (ARMA(p,q)).

4.2.2. Modelos autorregresivos AR(p)

El modelo autorregresivo de orden p es una aproximación al modelo de regresión lineal, en el que el valor presente se explica a partir de sus p valores pasados. Se define de la siguiente manera:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t$$

siendo c y ϕ constantes a determinar y Z_t un proceso de ruido blanco con varianza constante.

Aplicando el operador de retardo a la función anterior se obtiene:

$$\phi_p(B)X_t = Z_t$$

$\phi_p(B)$ es el polinomio autorregresivo y se define como:

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

Para que la serie sea estacionaria, las raíces de este polinomio deben ser mayores que la unidad en valor absoluto.

La función de autocorrelación simple de un modelo AR(p) decrece y tiende a cero rápidamente, mientras que la función de autocorrelación parcial tiene p valores significativos.

4.2.3. Modelos de medias móviles MA(q)

En un modelo de medias móviles de orden q, el valor presente se explica a partir de los q residuos pasados. Se define de la siguiente manera:

$$X_t = Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \dots - \theta_q Z_{t-q}$$

Aplicando el operador de retardo a la función anterior se obtiene:

$$X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)Z_t = \theta_q(B)Z_t$$

Este proceso siempre es estacionario, ya que está formado por la suma de varios procesos estacionarios. Además, si las raíces del polinomio anterior son mayores que la unidad en valor absoluto, el efecto de los valores pasados decrecerá con el tiempo y el proceso será invertible.

La función de autocorrelación simple de un modelo MA(q) tiene p valores significativos, mientras que la función de autocorrelación parcial decrece y tiende a cero rápidamente.

4.2.4. Modelos mixtos ARMA(p,q)

Los modelos mixtos son una combinación de los dos modelos anteriores. Se definen de la siguiente manera:

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \dots - \theta_q Z_{t-q}$$

Aplicando el operador de retardo a la función anterior se obtiene:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) Z_t$$

Este proceso es estacionario si lo es su parte autorregresiva y es invertible si lo es su media móvil.

Tanto la función de autocorrelación simple como la parcial de un modelo ARMA(p,q) disminuyen geométricamente y tienden a cero rápidamente.

4.2.5. Procesos integrados: Modelos ARIMA(p,d,q)

En la realidad, las series no suelen ser procesos estacionarios. Existen diferentes tipos de estacionariedad que toda serie debe cumplir:

- Estacionariedad en media: La serie oscila alrededor de un valor medio, es decir, su media es constante para toda la muestra. Por tanto, cuando una serie tiene tendencia creciente o decreciente, no será estacionaria en media.
- Estacionariedad en varianza: La dispersión de la serie a lo largo del tiempo es constante para toda la muestra. Por tanto, cuando una serie tiene una estacionalidad cada vez mayor, no será estacionaria en varianza.

Los procesos integrados no son estacionarios, ya que la relación de dependencia entre sus valores presentes y pasados va desapareciendo con el paso del tiempo. Aun así, pueden convertirse en estacionarios al diferenciarlos d veces en la parte regular, dando lugar a un proceso integrado de orden d. Se definen de la siguiente manera:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) Z_t$$

Al contrario de lo que ocurre con los modelos ARMA, tanto la función de autocorrelación simple como la parcial de un modelo integrado disminuyen linealmente y tienden a cero lentamente.

4.2.6. Modelos estacionales ARIMA(p,d,q)(P,D,Q)s

Las series con estacionalidad tampoco son procesos estacionarios, pero pueden convertirse en estacionarias al diferenciarlas s veces en la parte estacional, siendo s el periodo de la serie. Por tanto, una diferencia estacional es la diferente entre un valor y el anterior de su mismo periodo. Estos modelos se definen de la siguiente manera:

$$(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps})(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B^s)^D (1 - B)^d X_t = (1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs})(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) Z_t$$

4.2.7. Metodología Box-Jenkins

La metodología Box-Jenkins permite ajustar los modelos ARIMA a las series para su posterior predicción. Esta metodología consta de cuatro fases:

1. Identificación del modelo:

En esta etapa se debe seleccionar el modelo ARIMA que mejor reproduzca las características de la serie a partir de sus datos históricos. Para ello, es fundamental que la serie sea estacionaria tanto en media como en varianza, por lo que, en primer lugar, han de seleccionarse las transformaciones necesarias:

- Estacionariedad en media: diferencias en la parte regular y/o estacional entre los valores de la serie.
- Estacionariedad en varianza: transformaciones Box-Cox, como la raíz cuadrada o la inversa, aunque la más utilizada es la transformación logarítmica.

Una vez transformada la serie en estacionaria, se seleccionan provisionalmente los órdenes p y q de la parte regular y los órdenes P y Q de la parte estacional en caso de tenerla. Para ello, se utilizan las funciones de autocorrelación simple y parcial. Cuando es difícil identificar el modelo más adecuado, es preferible seleccionar varios posibles modelos y, posteriormente, elegir el mejor de ellos.

2. Estimación de los parámetros:

Tras haber identificado los posibles modelos, se estiman los parámetros de los mismos por máxima verosimilitud y se obtienen sus errores estándar y sus residuos. La máxima verosimilitud se basa en el cálculo de los residuos a partir de los valores de la serie estacionaria.

3. Diagnóstico del modelo:

En esta etapa se procede a evaluar la adecuación de los modelos estimados a los datos a través de diferentes contrastes:

- Significación estadística de los parámetros ($H_0: \varphi = 0$): Las estimaciones de los coeficientes del modelo deben ser significativas, por lo que aquellos parámetros no significativos serán eliminados del modelo, lo que supone la estimación de un modelo más sencillo. Además, hay que tener en cuenta las correlaciones entre los estimadores de los parámetros, ya que correlaciones altas pueden ser indicio de sobreparametrización.
- Análisis de los residuos: Los residuos del modelo deben ser un proceso de ruido blanco, es decir, deben tener media nula, varianza constante y retardos incorrelados. Este contraste puede realizarse a partir de los correlogramas simple y parcial; si todos los retardos se encuentran dentro de las bandas de confianza, los residuos estarán incorrelados.

4. Predicción de la serie:

El objetivo final de la modelización de las series consiste en predecir las mismas, por lo que, tras haber construido el modelo y haber comprobado su adecuación, será posible realizar las predicciones oportunas.

4.3. Análisis de tendencias

Existen diferentes metodologías que permiten analizar la tendencia de las series climáticas, pero la más utilizada por los investigadores es la prueba no paramétrica de Mann-Kendall (MK).

La presencia de autocorrelación en las series influye sobre los resultados de esta prueba, ya que puede causar problemas en la detección de la tendencia y en la evaluación del nivel de significación. Si la autocorrelación es positiva, aumenta la probabilidad de detección de tendencia, y si es negativa, disminuye. Para solventar este problema, Yue, et al. (2002b) plantean la técnica del preblanqueo libre de tendencia (trend-free prewhitening, TFPW).

Este análisis de tendencias será realizado con SAS Base.

4.3.1. Técnica del preblanqueo

La técnica del preblanqueo permite eliminar la correlación de las series, por lo que el primer paso será calcular el coeficiente de correlación de primer orden (r_1) de cada serie en caso de que exista.

Para detectar la presencia de autocorrelación en los residuos de una serie, la prueba más utilizada es la de Durbin-Watson, cuyo estadístico se define de la siguiente manera:

$$d = \frac{\sum_{t=2}^{t=T} (e_t e_{t-1})^2}{\sum_{t=1}^{t=T} e_t^2}$$

donde e_t es el residuo estimado para el periodo t y T es el número de observaciones.

Puede tener valores entre 0 y 4; un valor en torno a 2 indica la ausencia de correlación, valores entre 0 y 2 indican una correlación positiva y valores entre 2 y 4 indican una correlación negativa.

Las hipótesis asociadas a este contraste son las siguientes:

- H_0 : No existe correlación de primer orden.
- H_1 : Existe correlación de primer orden.

Si el coeficiente de correlación de primer orden no es significativamente distinto de cero, la serie se considera independiente y la prueba MK se aplica sobre la serie original, pero si r_1 es significativamente distinto de cero, la serie original ha de ser preblanqueada para eliminar la correlación.

En el caso de existir correlación en la serie, la expresión utilizada para preblanquear la misma es la siguiente:

$$y_t = \frac{y_t - r_1 y_{t-1}}{1 - r_1}$$

Las series preblanqueadas tendrán la misma tendencia que la serie original, aunque la varianza de sus residuos será mayor.

4.3.2. Contrastes de Mann-Kendall y Theil-Sen

Una vez eliminada la correlación de las series, es posible aplicar el contraste de Mann-Kendall. Indica la significancia estadística de las tendencias, es decir, si una serie tiene o no tendencia, así como si esta es positiva o negativa, pero este estadístico no aporta el valor de su pendiente. Por tanto, la hipótesis nula asociada a esta prueba es que no existe una tendencia monotónica en la serie.

Se expresa como el número de diferencias positivas menos el número de diferencias negativas:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k)$$

Si S es un número positivo, las observaciones obtenidas posteriormente en el tiempo tienden a ser más grandes que las observaciones anteriores, es decir, la tendencia es positiva, y si S es un número negativo, las observaciones obtenidas posteriormente en el tiempo tienden a ser más pequeñas que las observaciones anteriores.

Para poder estimar el valor de la pendiente de la tendencia se emplea el contraste de Theil-Sen, que es la mediana de las pendientes entre cada par de puntos dentro de un conjunto de datos.

4.3.3. Regresión lineal

El modelo de regresión lineal permite analizar la posible relación lineal existente entre una variable dependiente (Y) y un conjunto de variables independientes (X), así como predecir los valores que tomará esa variable dependiente en el futuro. Se expresa de la siguiente manera:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon$$

ε es la perturbación aleatoria y representa el error cometido, es decir, la parte de la variable dependiente no explicada a partir de las variables independientes. β constituye el estimador de cada una de las variables independientes y mide la influencia que tienen estas sobre la variable dependiente, siendo β_0 el término constante.

Para poder llevar a cabo la predicción, es necesario estimar el valor de esos parámetros que minimiza el error cometido por el modelo. El estimador de mínimos cuadrados es aquel que minimiza la suma de cuadrados de los errores, cuya expresión es la siguiente:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Se debe tener en cuenta que, a la hora de realizar la predicción, no se pueden incluir dos o más variables independientes continuas que estén muy correlacionadas, ya que no

sería posible invertir la matriz $X'X$. Además, el número de parámetros incluidos en el modelo debe ser muy inferior al número de observaciones con el fin de evitar problemas en la estimación y de sobreajuste.

5. DESARROLLO DEL TRABAJO Y RESULTADOS

5.1. Homogeneización de las series

Como ya se ha comentado anteriormente, los datos de los que se dispone abarcan dos periodos de tiempo; el primero, desde el 1 de enero de 1920 hasta el 29 de febrero de 2012 y, el segundo, desde el 7 de mayo de 2013 hasta el 31 de mayo de 2017. Una de las condiciones necesarias para que el paquete de R *Climatol* sea capaz de realizar el procedimiento de homogeneización de las series climáticas consiste en que no debe existir ningún momento en el tiempo en el no se disponga de datos en ninguna estación meteorológica. Como los datos disponibles no cubren un periodo temporal completo, este procedimiento ha de realizarse dos veces para cada una de las variables climáticas. Además, cada periodo de tiempo debe tener datos que abarquen años completos, lo que supone la pérdida de los dos meses de 2012 en el primer caso, y de los meses de 2013 y de 2017 en el segundo caso.

Otro dato a tener en cuenta consiste en que un número muy reducido de estaciones cuenta con datos durante las primeras décadas del periodo 1920-2011, lo cual supondría que la homogeneización y el rellenado de datos ausentes en esos años fuera muy poco fiable. Por tanto, se decide eliminar los 30 primeros años, con lo que quedaría un periodo de 62 años seguidos.

En definitiva, los dos periodos de tiempo que, finalmente, se van a tratar en este estudio son enero de 1950 a diciembre de 2011 y enero de 2014 a diciembre de 2017.

5.1.1. Preparación de los ficheros de entrada

Para realizar el proceso de homogeneización, *Climatol* necesita dos ficheros de entrada, uno con las estaciones y otro con los datos de la variable climática a estudiar.

El fichero de estaciones, cuya extensión debe ser “.est”, está formado por el código, el nombre y la altitud en metros de cada estación, así como por su longitud y latitud en grados con decimales y con el signo adecuado para indicar Norte, Sur, Este u Oeste. Estas coordenadas se encuentran en grados, minutos y segundos, por lo que han de ser transformadas a grados con decimales de la siguiente manera:

$$\text{Grados decimales} = \text{Grados} + \frac{\text{Minutos}}{60} + \frac{\text{Segundos}}{3600}$$

El fichero de datos, cuya extensión debe ser “.dat”, contiene únicamente los valores de la variable a estudiar, ordenados por orden cronológico desde la primera estación hasta la última. Este fichero carece de referencia temporal, por lo que todos los datos deben estar presentes para todo el periodo de estudio, representando los valores ausentes, por ejemplo, con “NA”.

Para ayudar en la preparación de los ficheros de entrada con este formato, R provee la función “db2dat”, que genera ambos ficheros extrayendo las series de una base de datos a través de una conexión ODBC (Open DataBase Connectivity). Para ello, se va a utilizar MySQL, un sistema de gestión de bases de datos relacional, ya que dispone de dicha conexión. Una vez importados los datos en el mismo, se carga el paquete de R “RODBC”. Es necesario conocer tanto el nombre de la base de datos creada en MySQL como los nombres de las tablas y campos en los que están almacenados los datos diarios de la variable a estudiar y sus fechas correspondientes, así como los de las coordenadas, códigos y nombres de las diferentes estaciones.

En el primer periodo, se ha decidido no rechazar ninguna serie, ya que todas ellas tienen al menos 24 años de datos en total, lo cual se considera suficiente, pero en el segundo periodo, se han rechazado las series con menos de 1 año de datos, las cuales ascienden a un total de 23, quedando 693 estaciones válidas.

5.1.2. Primer análisis exploratorio de los datos diarios originales para el control de calidad

Los umbrales para el rechazo de datos atípicos y la detección del punto de inflexión de las inhomogeneidades pueden variar en gran medida en función de la periodicidad de los datos y las correlaciones cruzadas de las series. Por tanto, se va a realizar un primer análisis exploratorio sobre los datos originales para un control de calidad confiable, con lo que se obtienen diferentes gráficos de diagnóstico.

A continuación, se muestran las correlaciones entre las series y su clasificación en grupos con variabilidad similar representados en un mapa (se han seleccionado los gráficos de las precipitaciones del primer periodo y los de la temperatura máxima del segundo periodo a modo de ejemplo). Las correlaciones son generalmente más bajas cuando la distancia entre las estaciones es mayor. Cuanto más altas sean estas correlaciones, mayor será la fiabilidad de la homogeneización y el relleno de datos ausentes. Deben ser positivas, aunque algunas de ellas son negativas, probablemente debido a la existencia de discontinuidades geográficas que produzcan diferencias climáticas. En cuanto al mapa de estaciones, en el primer periodo se agrupan las mismas en un total de 8 clusters para las temperaturas máximas, y en 7 clusters, para las lluvias. En el segundo periodo se ve mucho más clara la distribución de los grupos al estar activos muchos más observatorios. Como cabría esperar, estos grupos están formados por estaciones situadas aproximadamente en la misma zona.

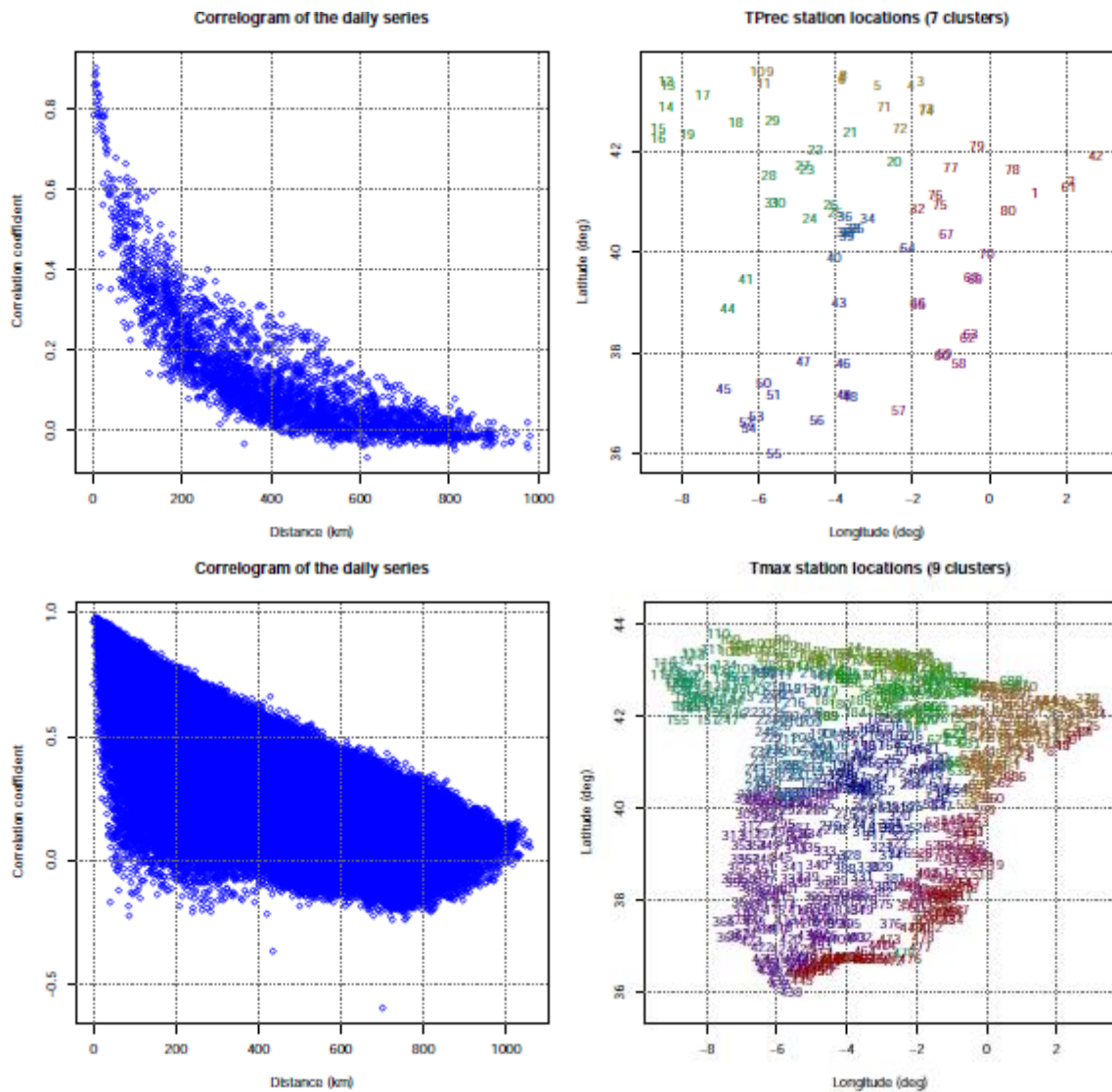


Figura 3. Correlaciones entre las series (izquierda) y mapa de las estaciones (derecha; los colores identifican grupos de estaciones con variabilidad similar)

Los siguientes gráficos muestran algunas de las anomalías de las series originales. Se han seleccionado dos gráficos del segundo periodo para la temperatura máxima a modo de ejemplo. La serie de la izquierda parece bastante homogénea, pero la serie de la derecha es claramente heterogénea, ya que se produce un cambio brusco en la media, alcanzando valores de SNHT máximos muy elevados (288 y 489) en el mismo punto.

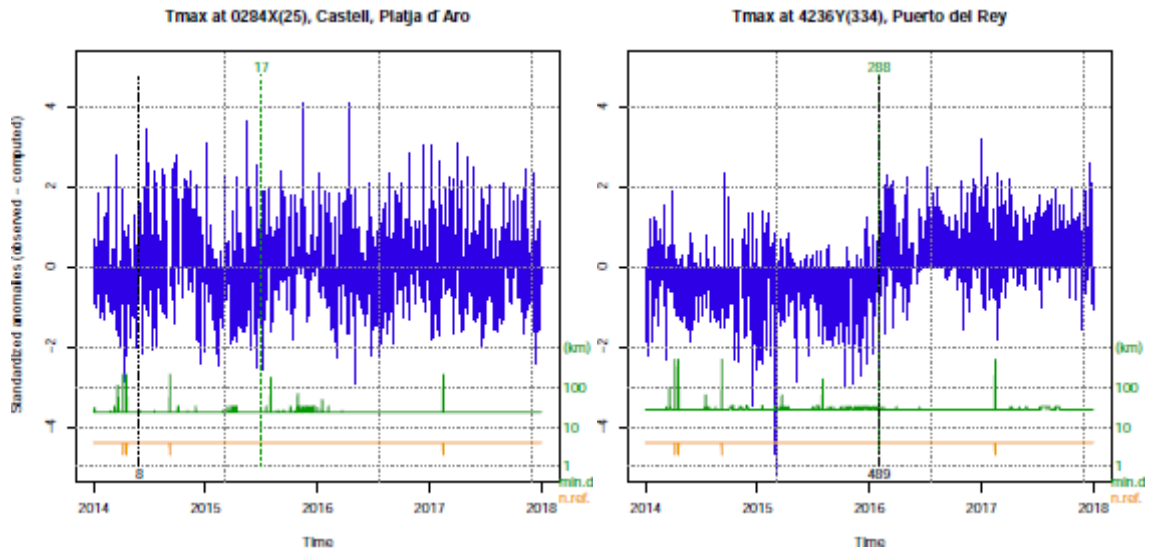


Figura 4. Anomalías de una serie homogénea (izquierda) y una heterogénea (derecha)

Los histogramas de anomalías estandarizadas (figura 5) permiten elegir el umbral más adecuado para rechazar los datos muy anómalos, suponiendo que son errores y pueden eliminarse. Tan solo en dos casos (temperatura máxima y precipitaciones del segundo periodo) los histogramas no muestran ningún tipo de sesgo, por lo que podrían aceptarse todos los datos, pero en el resto de los casos, los histogramas se encuentran desplazados en mayor o menor medida hacia la izquierda o derecha, por lo que algunos datos son considerados anómalos. A continuación, se muestran dos de ellos (temperatura máxima y mínima del segundo periodo):

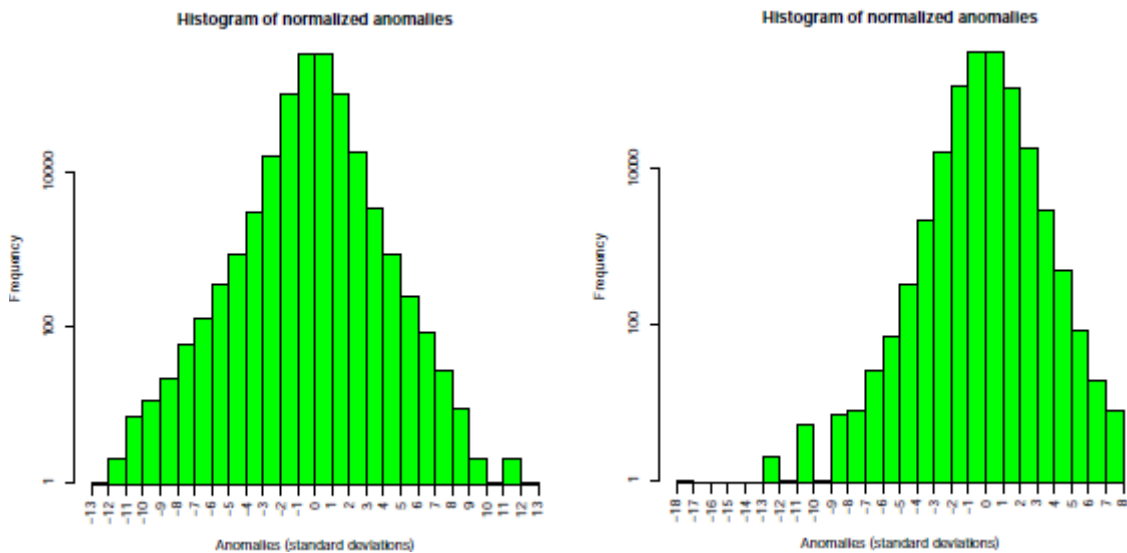


Figura 5. Histogramas de anomalías estandarizadas

Por último, se muestra un gráfico que indica la calidad o singularidad de las series finales (figura 6). Muestra el error típico (RMSE, por sus siglas en inglés) final de cada una de las estaciones y sus valores de SNHT. Los errores se calculan al comparar los datos estimados y los observados en cada serie. Un valor alto puede indicar una mala calidad, pero también podría deberse a que la estación se encuentre en un sitio peculiar con un microclima distinto. Las series homogéneas de estaciones que comparten el clima común de una región tenderán a agruparse en la parte inferior izquierda del gráfico. A

continuación, se muestra este gráfico para las precipitaciones del primer periodo y para la temperatura máxima del segundo periodo.

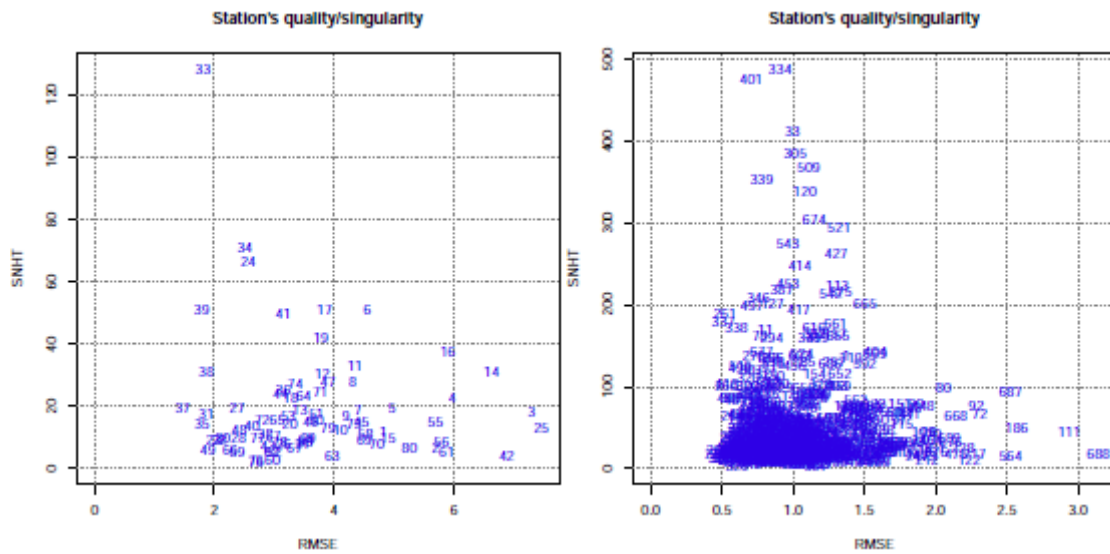


Figura 6. Gráfico de la calidad/singularidad de las series finales

5.1.3. Agregación de los datos diarios en series mensuales

Después de todas estas consideraciones, se podrían homogeneizar las series, pero, puesto que las series diarias muestran una alta variabilidad que reduce la eficiencia de la detección de sus inhomogeneidades, es preferible agregarlas y homogeneizar primero las series mensuales. Para las series de temperatura, los datos mensuales serán el promedio, pero para las series de precipitaciones, se calcularán los totales mensuales.

5.1.4. Homogeneización de las series mensuales

Una vez obtenidas las series mensuales a partir de los valores diarios, se procede a homogeneizar las mismas. En esta fase se han de inspeccionar los histogramas de valores máximos de SNHT para las etapas de detección 1 (en ventanas escalonadas superpuestas) y 2 (en las series completas). Permiten identificar los valores de SNHT más adecuados, siendo 25 el valor predeterminado.

En el primer periodo, este valor parece apropiado para las precipitaciones mensuales en ambas etapas, ya que en ese punto se produce un mínimo. Sin embargo, en el caso de las temperaturas, se han buscado otros valores más adecuados, los cuales coinciden tanto en las máximas como en las mínimas y medias. En la primera etapa, parece que el valor de SNHT más apropiado es 60 y, en la segunda, podría ser 90. Por otro lado, en el segundo periodo no se producen inhomogeneidades en la primera etapa, pero en el caso de las series completas, un SNHT de 25 parece suficiente para las temperaturas, aunque este se reduce a 16 para las precipitaciones.

En aquellos casos en los que se haya cambiado el valor de SNHT, se vuelven a homogeneizar las series mensuales modificando este parámetro.

A continuación, se muestran los histogramas de valores máximos de SNHT para ambas etapas de las precipitaciones del primer periodo (figura 7), así como los histogramas de las series completas de la temperatura máxima y precipitaciones del segundo periodo (figura 8).

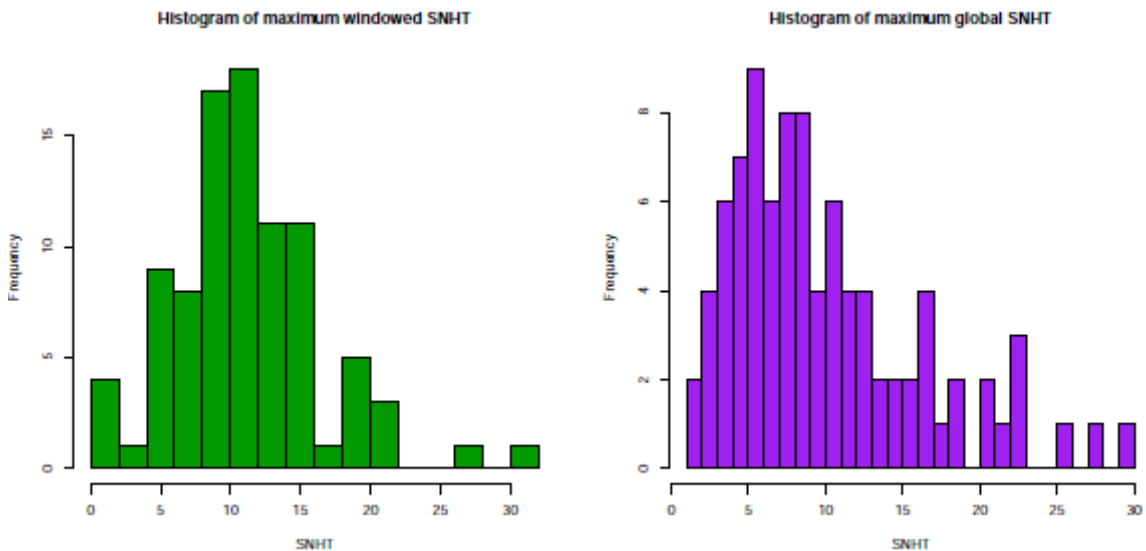


Figura 7. Histogramas de valores máximos de SNHT encontrados en ventanas escalonadas superpuestas (izquierda) y en las series completas (derecha)

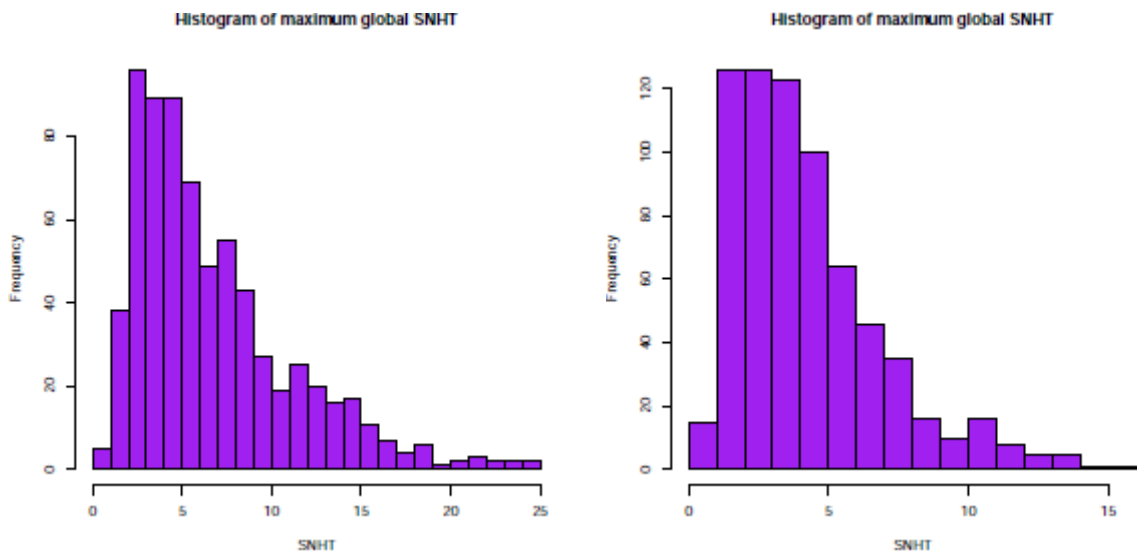


Figura 8. Histogramas de valores máximos de SNHT encontrados en las series completas

Por último, queda por conocer el número total de inhomogeneidades existentes en cada una de las series y variables. En la primera etapa del periodo 1950-2011 se han detectado 8, 22, 6 y 4 cambios en la media en la temperatura máxima, mínima y media y precipitaciones, respectivamente, y en la segunda etapa se han producido 17, 35, 25 y 7 cambios. En cuanto al periodo 2014-2017, en la primera etapa no se han detectado inhomogeneidades, pero en el caso de las series completas se han producido 10, 7, 12 y 6 cambios en la media en cada una de las variables.

En el anexo se encuentra detallado el momento en el que se han producido estas inhomogeneidades, así como sus respectivos valores SNHT máximos.

A modo de ejemplo, la figura 9 muestra una misma estación meteorológica en la que se han producido dos cambios en la media con SNHT máximos de 87 y 62. Esta serie se divide en tres fragmentos homogéneos para su posterior reconstrucción.

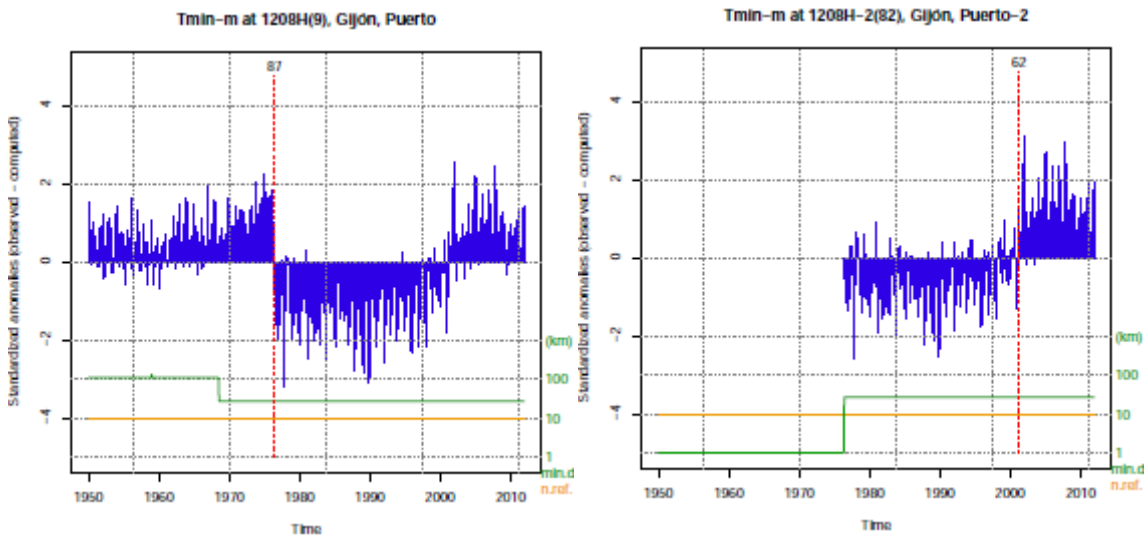


Figura 9. Inhomogeneidades de una serie detectadas en la etapa 1

A continuación, se muestra el gráfico de reconstrucción de la serie anterior a partir de cada subperiodo homogéneo. Los datos originales aparecen en negro, y en la parte inferior se muestran las correcciones aplicadas, las cuales presentan variaciones estacionales.

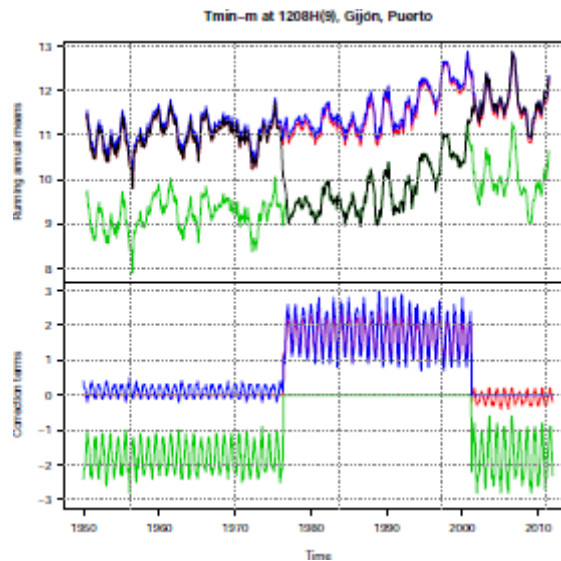


Figura 10. Reconstrucción de la serie a partir de los tres subconjuntos homogéneos

5.1.5. Homogeneización de las series diarias con los puntos de corte mensuales

Los puntos de corte detectados a partir de la homogeneización de las series mensuales van a permitir el ajuste de los datos diarios iniciales. El corte se efectuará el día 1 del

mes en el que se produce el salto. Además, en esta fase ha de establecerse el umbral de rechazo de valores atípicos detectado en el control de calidad de los datos diarios.

Por cada cambio en la media producido en cada estación se creará una nueva serie, por lo que se seleccionará únicamente aquella que contenga el mayor número de datos reales iniciales para continuar el estudio.

De esta forma, se obtiene un conjunto de datos homogeneizado, sin valores ausentes y con los datos atípicos corregidos, con un total de 1.268.120 observaciones en el periodo 1950-2011 y 81.816 en el periodo 2014-2017.

5.2. Análisis de series temporales

El siguiente objetivo de este estudio consiste en analizar la evolución que han tenido en el tiempo la temperatura máxima, mínima y media y las precipitaciones desde 1950 hasta 2011, tratando de detectar el modelo que mejor se ajuste a cada una de ellas. Posteriormente, se tratará de predecir su comportamiento hasta el año 2017. Los datos del periodo 2014-2017 van a permitir comprobar la fiabilidad de esta predicción comparando los datos estimados con los reales. Para ello, se van a seleccionar únicamente las estaciones meteorológicas de las que se tenga información en ambos periodos, las cuales ascienden a un total de 56. En algunos casos, el identificador de estas estaciones no coincide en ambos periodos y, en otros casos, estaciones con el mismo identificador no son la misma, ya que sus coordenadas y su altitud no coinciden.

La siguiente tabla recoge cada una de estas estaciones con el identificador que tienen en cada periodo, así como la provincia a la que pertenecen.

ID PERIODO 1	ID PERIODO 2	PROVINCIA	ID PERIODO 1	ID PERIODO 2	PROVINCIA
0016A		Tarragona	3195		Madrid
0076		Barcelona	3260B		Toledo
0367		Girona	3469A		Cáceres
1082		Bizkaia	4121		Ciudad Real
1109		Cantabria	4452		Badajoz
1111	1111X	Cantabria	4642E		Huelva
1212E		Asturias	5270B		Jaén
1249I	1249X	Asturias	5402		Córdoba
1387		A Coruña	5530E		Granada
1387E		A Coruña	5783		Sevilla
1428		A Coruña	5796		Sevilla
1484C		Pontevedra	5960		Cádiz
1495		Pontevedra	5973		Cádiz
1505		Lugo	6001		Cádiz
1549		León	6325O		Almería
1690A		Ourense	7178I		Murcia
2030		Soria	8019		Alicante
2331	2331X	Burgos	8025		Alicante
2422		Valladolid	8096		Cuenca
2444		Ávila	8178D		Albacete

2462	Madrid	8368U		Teruel
2465	Segovia	8416	8416Y	Valencia
2614	Zamora	8500A		Castellón
2867	Salamanca	9170		La Rioja
2870	Salamanca	9263D		Navarra
3013	Guadalajara	9381I		Teruel
3129	Madrid	9390		Zaragoza
3191E	Madrid	9898		Huesca

Tabla 2. Estaciones meteorológicas operativas en ambos periodos

Para realizar la predicción de las series, se ha de encontrar el modelo que mejor se ajuste a cada una de ellas. Se va a buscar un único modelo para cada variable de tal modo que sirva para todas las estaciones meteorológicas. Se van a escoger tres estaciones localizadas en lugares diferentes de la Península con el fin de comprobar que el modelo más adecuado para cada una de ellas es el mismo: 3195 (Madrid), 0076 (Barcelona) y 4642E (Huelva). Aun así, solo se mostrarán los resultados de la estación de Madrid.

Los datos diarios van a ser agregados a series mensuales, por lo que cada una de ellas va a estar formada por 744 observaciones. En el caso de las temperaturas, los datos mensuales constituirán la media de los datos diarios, pero en el caso de las precipitaciones, serán el total acumulado de cada mes.

5.2.1. Representación gráfica y descomposición estacional

A continuación, se representan gráficamente las series de temperatura máxima, mínima y media y precipitaciones totales de la estación de Madrid, así como su descomposición estacional:

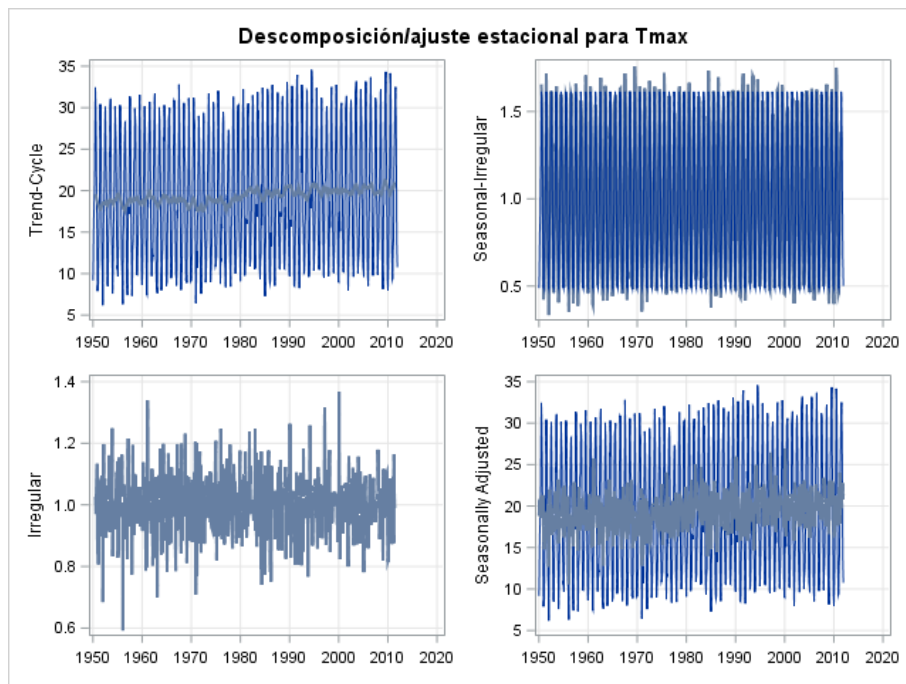


Figura 11. Representación gráfica y descomposición estacional de la temperatura máxima

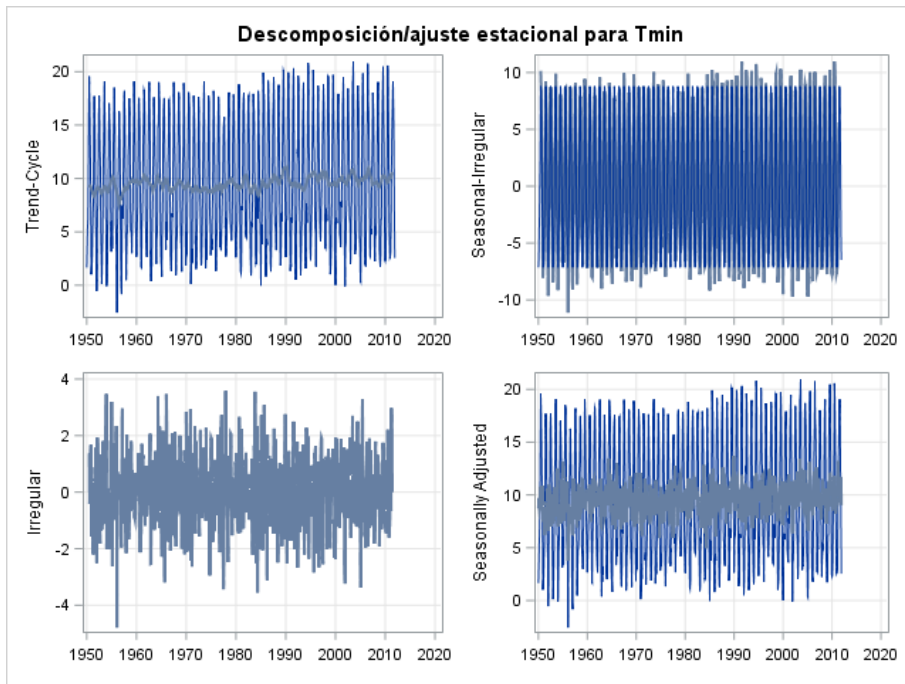


Figura 12. Representación gráfica y descomposición estacional de la temperatura mínima

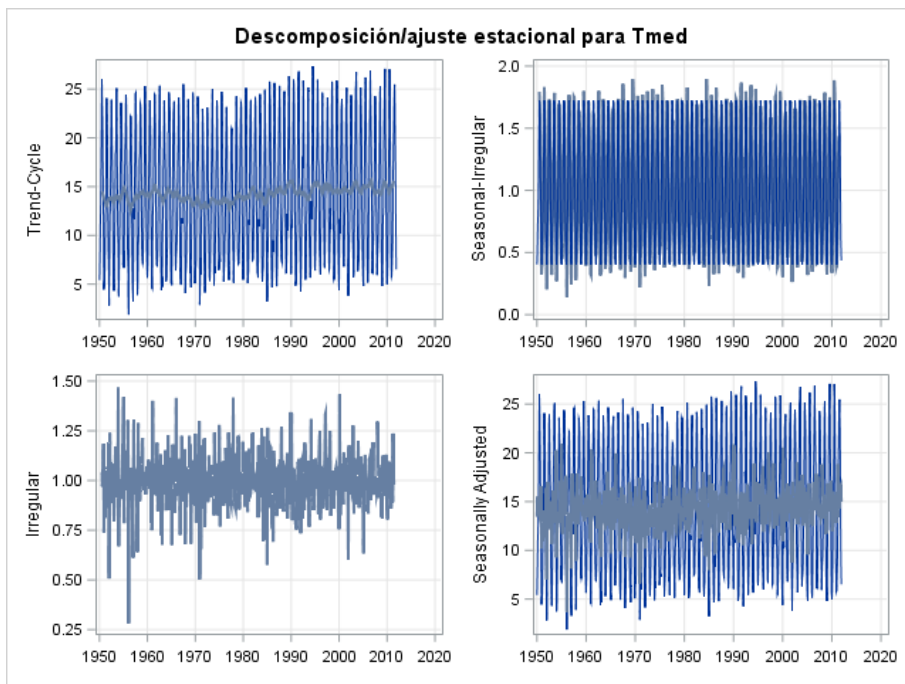


Figura 13. Representación gráfica y descomposición estacional de la temperatura media

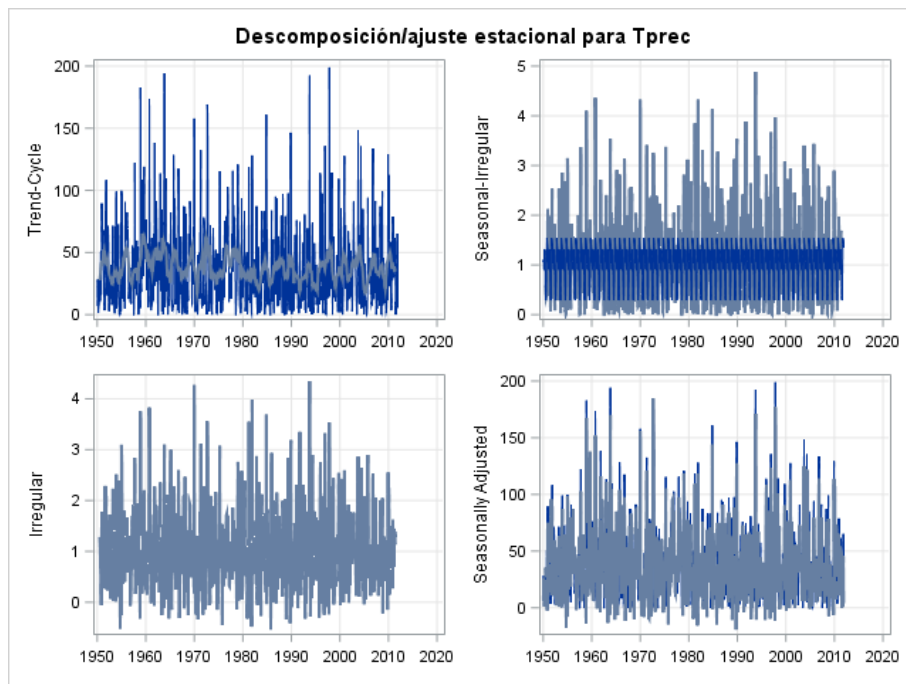


Figura 14. Representación gráfica y descomposición estacional de las precipitaciones

Las tres series de temperatura siguen el mismo patrón; se observa un leve incremento de las mismas con el paso del tiempo y con repeticiones cada 12 meses. Esta tendencia es regular, ya que las series tienen aproximadamente la misma pendiente en todos los puntos. Por tanto, no oscilan alrededor de un valor medio, es decir, las medias no son las mismas para toda la muestra, por lo que no parece que sean estacionarias en media. La estacionalidad de estas series también es regular, ya que su dispersión se mantiene constante para toda la muestra, por lo que sí que parecen ser estacionarias en varianza.

En cuanto a las precipitaciones totales, no se distingue ningún tipo de tendencia, pero podría existir un componente estacional de 12 meses, aunque los valores atípicos dificultan su detección. Además, esta serie parece ser estacionaria tanto en media como en varianza.

Como las tres series de temperatura tienen una distribución muy similar, se va a explicar únicamente la temperatura máxima, pero se mostrarán los resultados finales de todas ellas.

5.2.2. Análisis de la estacionariedad de las series y ajuste de los modelos ARIMA

Anteriormente, se ha tratado de detectar la estacionariedad de las series; parece que las temperaturas son estacionarias en varianza, pero no en media, y las precipitaciones parecen ser estacionarias tanto en media como en varianza, lo cual se va a comprobar a continuación a partir de las funciones de autocorrelación de las series.

En primer lugar, se va a analizar la estacionariedad de las temperaturas. El correlograma simple (ACF) no decrece hacia 0 conforme aumenta el número de retardos, por lo que no es una serie estacionaria.

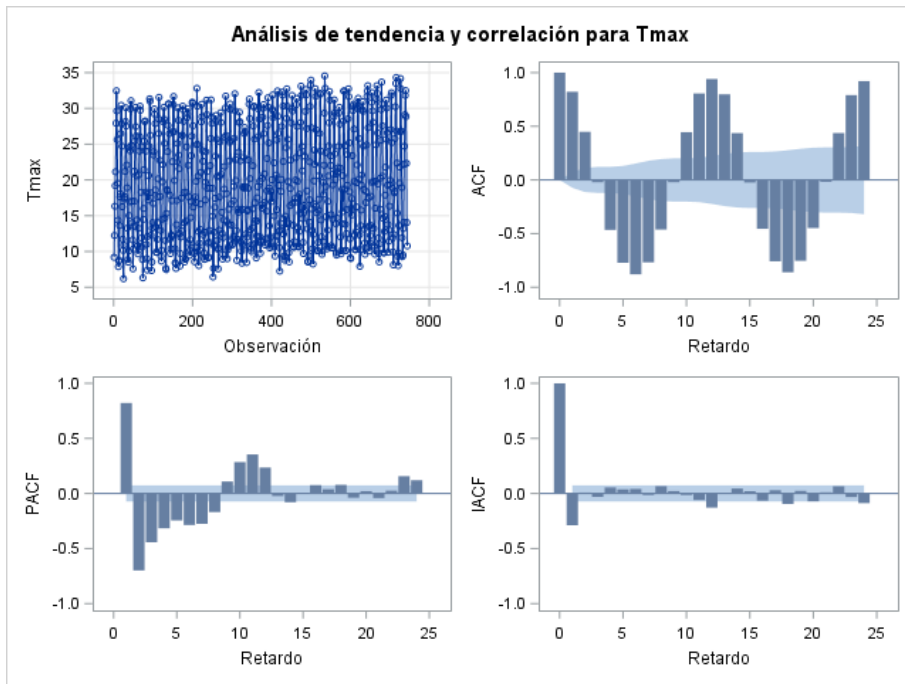


Figura 15. Funciones de autocorrelación simple y parcial de la temperatura máxima

Para transformarla en estacionaria, hay que identificar si se debe diferenciar la parte regular y/o la estacional de la serie. Para ello, se vuelven a representar los correlogramas diferenciando la parte regular.

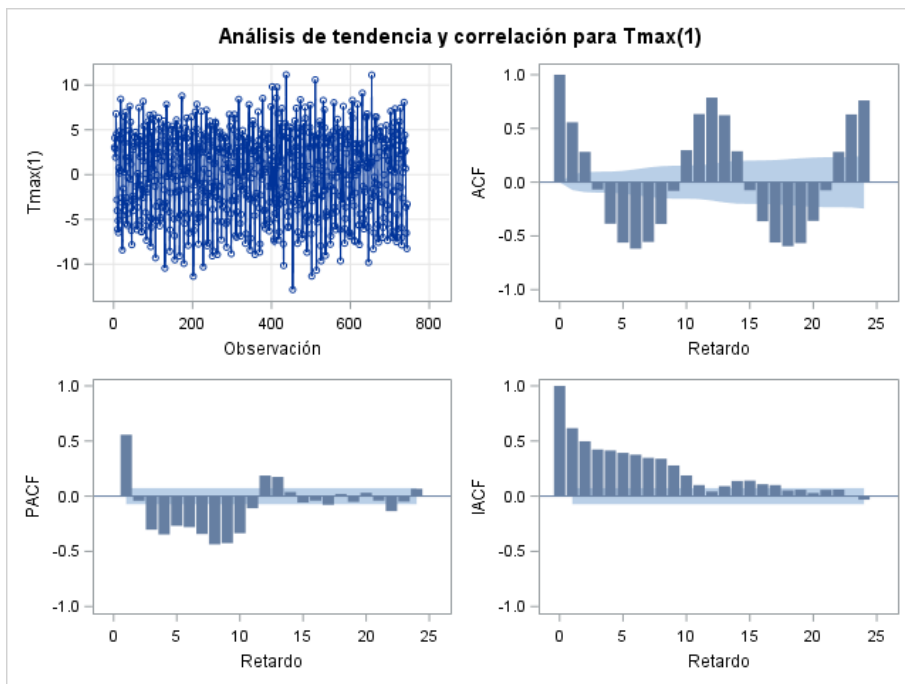


Figura 16. Funciones de autocorrelación simple y parcial de la temperatura máxima con una diferencia en la parte regular

La serie continúa sin ser estacionaria, ya que el correlograma simple sigue teniendo la misma estructura, por lo que se va a diferenciar únicamente la parte estacional.

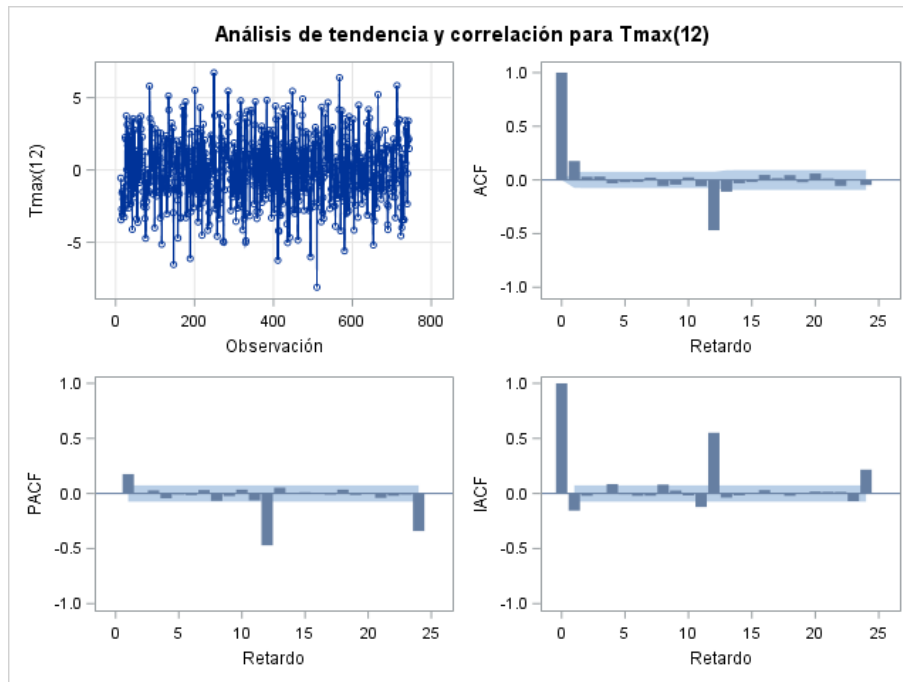


Figura 17. Funciones de autocorrelación simple y parcial de la temperatura máxima con una diferencia en la parte estacional

En este caso, el correlograma simple se anula a partir del segundo retardo y la serie oscila alrededor de un valor medio, por lo que ya es estacionaria tanto en media como en varianza.

Una vez transformada la serie en estacionaria, se va a tratar de identificar el modelo ARIMA más adecuado a través de los correlogramas anteriores. En la parte regular el único retardo significativo tanto en las autocorrelaciones simples como en las parciales es el primero, lo cual puede indicar que la serie tenga un autorregresivo o una media móvil de orden 1. En la parte estacional, únicamente es significativo el primer retardo (12) en las autocorrelaciones simples, pero en el caso de las parciales, los retardos decrecen tendiendo a 0, por lo que se podría tratar de una media móvil.

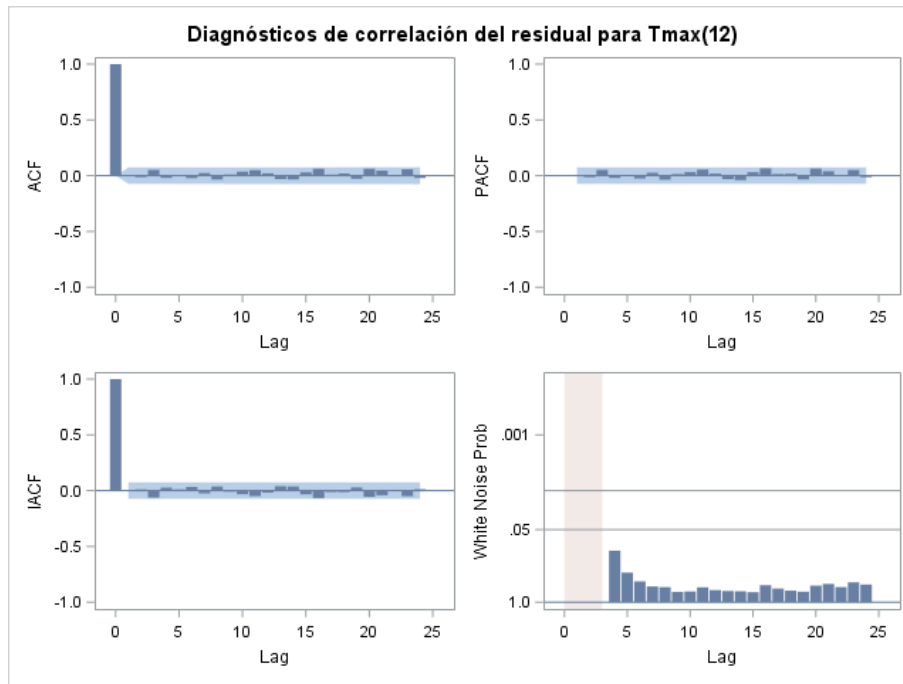


Figura 18. Funciones de autocorrelación simple y parcial de la temperatura máxima con un autorregresivo y una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional

Todos los retardos de ambos correlogramas se anulan, pero el valor estimado tanto de la media como del autorregresivo y la media móvil de la parte regular no son significativos (tabla 11), es decir, se acepta la hipótesis nula de que esos valores son iguales a 0, por lo que deben ser eliminados del modelo.

Estimación por mínimos cuadrados condicional					
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t	Retardo
MU	0.02019	0.01191	1.69	0.0906	0
MA1,1	0.03857	0.19835	0.19	0.8459	1
MA2,1	0.86186	0.01940	44.43	<.0001	12
AR1,1	0.22488	0.19363	1.16	0.2459	1

Tabla 3. Estimaciones y significatividad de los parámetros del modelo ARIMA ajustado

En el caso del autorregresivo y la media móvil el problema podría encontrarse en que ambos están muy correlacionados (tabla 12), por lo que se va a eliminar únicamente la media móvil, ya que su valor estimado es mucho más próximo a 0.

Correlaciones de las estimaciones de parámetro				
Parámetro	MU	MA1,1	MA2,1	AR1,1
MU	1.000	-0.003	-0.092	-0.004
MA1,1	-0.003	1.000	0.048	0.982
MA2,1	-0.092	0.048	1.000	0.065
AR1,1	-0.004	0.982	0.065	1.000

Tabla 4. Correlaciones entre los parámetros del modelo ARIMA ajustado

Tras eliminar la constante y la media móvil de la parte regular, el autorregresivo pasa a ser significativo (tabla 13) y todos los retardos de los correlogramas se anulan (figura 19).

Estimación por mínimos cuadrados condicional					
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t	Retardo
MA1,1	0.85419	0.01984	43.04	<.0001	12
AR1,1	0.19055	0.03654	5.21	<.0001	1

Tabla 5. Estimaciones y significatividad de los parámetros del modelo ARIMA ajustado

A partir de los correlogramas (figura 19) y de la tabla de contrastes de aleatoriedad de los residuos (tabla 14) se verifica que en todos ellos se acepta la hipótesis nula de que sean aleatorios, es decir, no están correlacionados, por lo que el modelo explica toda la estructura de dependencia de la serie y esta sigue un proceso de ruido blanco.

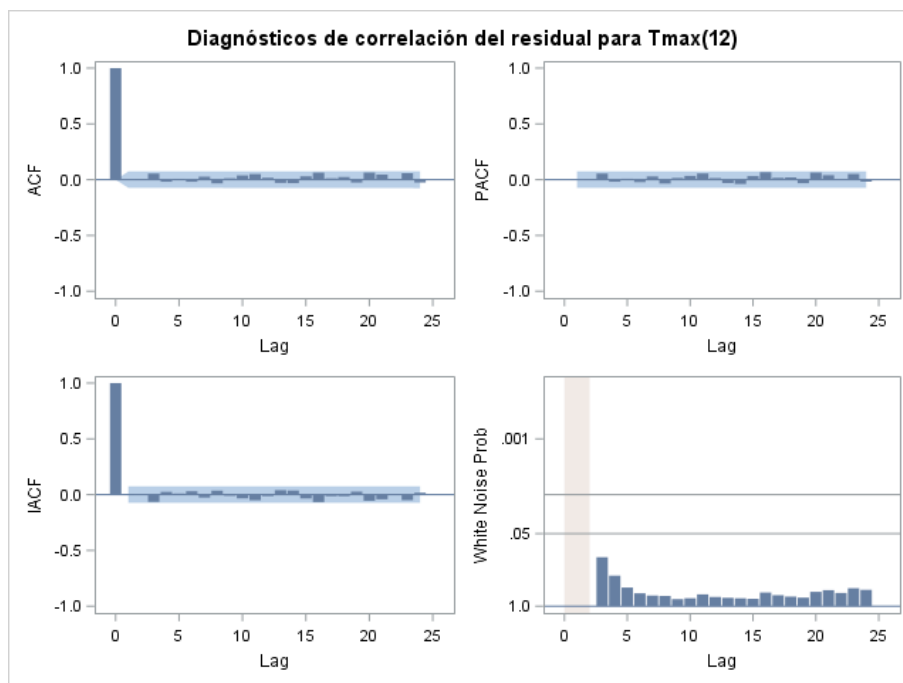


Figura 19. Funciones de autocorrelación simple y parcial de la temperatura máxima con una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional

Comprobación de autocorrelación de residuales										
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones						
6	2.86	4	0.5822	-0.001	-0.004	0.056	-0.018	0.009	-0.018	
12	7.66	10	0.6616	0.028	-0.032	0.013	0.039	0.051	0.018	
18	13.35	16	0.6468	-0.027	-0.030	0.031	0.065	0.013	0.025	
24	21.79	22	0.4722	-0.024	0.065	0.047	-0.003	0.060	-0.024	
30	29.15	28	0.4048	-0.016	-0.038	0.057	0.057	0.021	-0.033	
36	37.81	34	0.2996	0.014	0.035	0.060	0.075	0.012	0.021	
42	39.53	40	0.4915	0.020	0.026	0.022	0.011	-0.022	-0.008	
48	42.21	46	0.6315	0.023	-0.004	-0.000	-0.039	0.012	0.035	

Tabla 6. Contrastes de aleatoriedad de los residuos del modelo ARIMA ajustado

Además, los diagramas de normalidad (figura 20) muestran que los residuos se ajustan a una distribución normal.

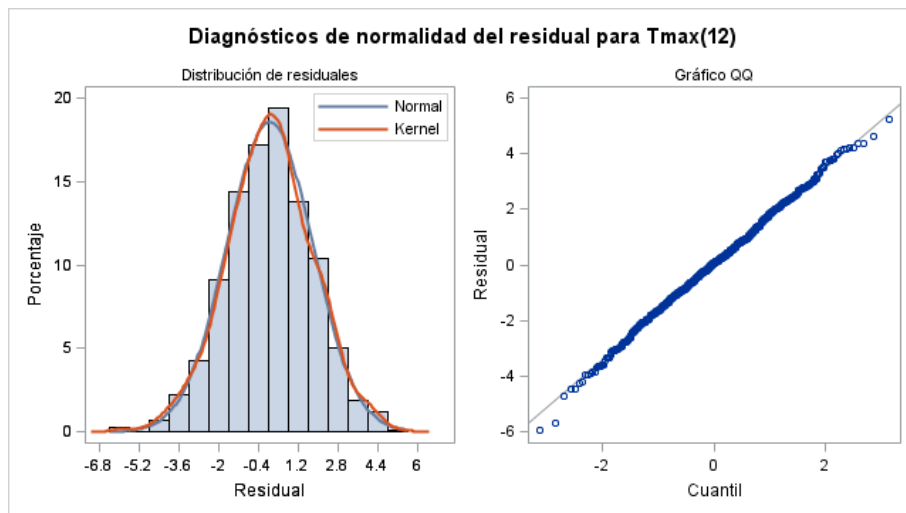


Figura 20. Diagramas de normalidad de los residuos de la temperatura máxima

A continuación, se muestran los resultados de las series de temperatura mínima y media:

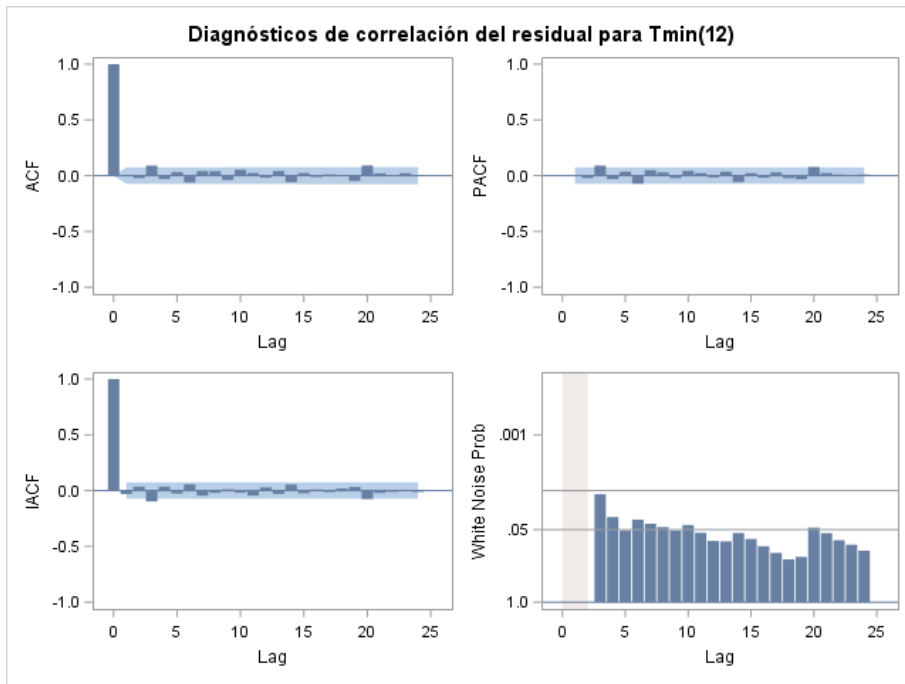


Figura 21. Funciones de autocorrelación simple y parcial de la temperatura mínima con una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional

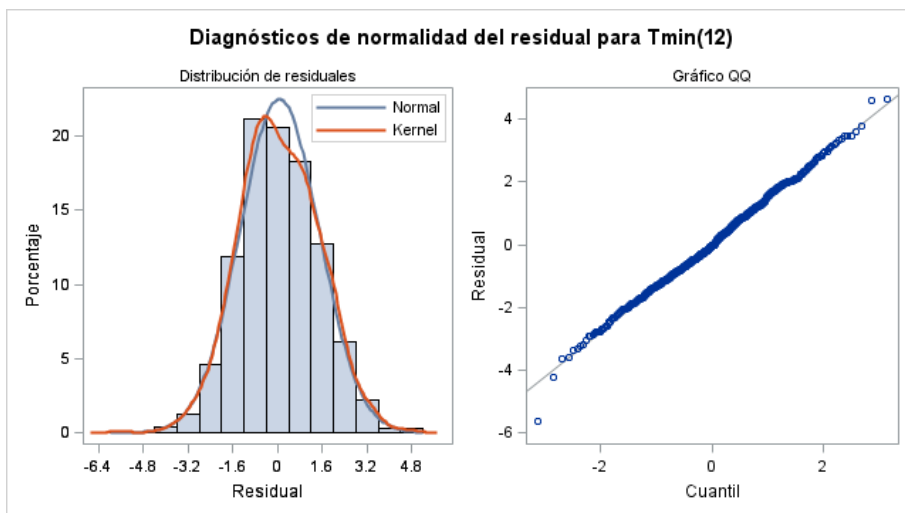


Figura 22. Diagramas de normalidad de los residuos de la temperatura mínima

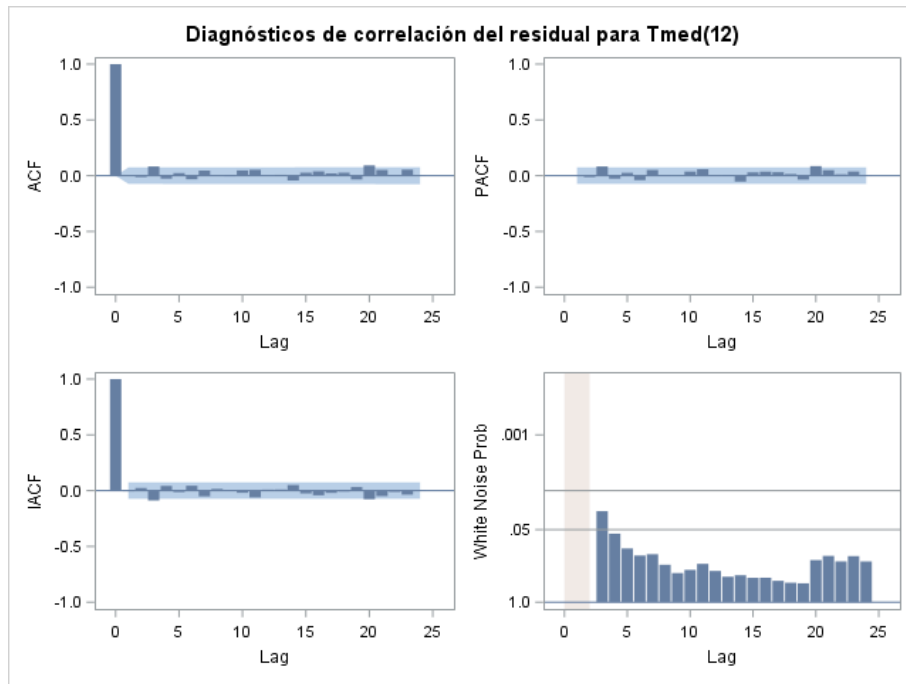


Figura 23. Funciones de autocorrelación simple y parcial de la temperatura media con una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional

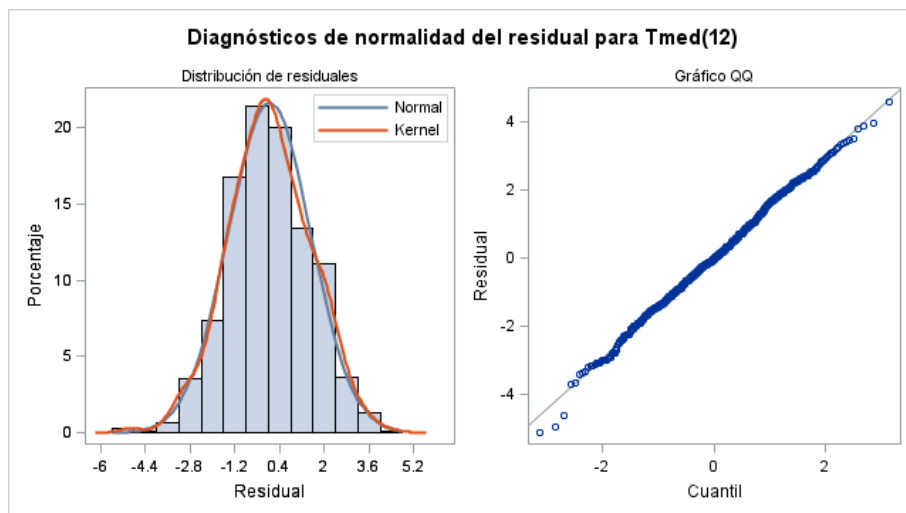


Figura 24. Diagramas de normalidad de los residuos de la temperatura media

Por tanto, el modelo ARIMA más adecuado para realizar la predicción de las series de temperatura máxima, mínima y media va a ser un modelo diferenciado en la parte estacional con un autorregresivo en la parte regular y una media móvil en la parte estacional: ARIMA (1,0,0) (0,1,1)₁₂.

Los parámetros estimados de cada uno de los modelos ajustados son los siguientes:

Factores autoregresivos		Factores autoregresivos		Factores autoregresivos	
Factor 1:	1 - 0.19055 B**(1)	Factor 1:	1 - 0.19849 B**(1)	Factor 1:	1 - 0.19565 B**(1)
Factores de la media móvil		Factores de la media móvil		Factores de la media móvil	
Factor 1:	1 - 0.85419 B**(12)	Factor 1:	1 - 0.88252 B**(12)	Factor 1:	1 - 0.85557 B**(12)

Tabla 7. Parámetros estimados de los modelos ajustados para la temperatura máxima, mínima y media, respectivamente

Con ellos, se obtienen las expresiones algebraicas de estos modelos ajustados. La expresión correspondiente a la temperatura máxima es la siguiente:

$$(1 - \phi_1 B)(1 - B^{12})X_t = (1 - \theta_1 B^{12})Z_t$$

$$(1 - 0,19055B)(1 - B^{12})X_t = (1 - 0,85419B^{12})Z_t$$

$$(1 - 0,19055B - B^{12} + 0,19055B^{13})X_t = (1 - 0,85419B^{12})Z_t$$

$$X_t - 0,19055X_{t-1} - X_{t-12} + 0,19055X_{t-13} = Z_t - 0,85419Z_{t-12}$$

$$X_t = 0,19055X_{t-1} + X_{t-12} - 0,19055X_{t-13} + Z_t - 0,85419Z_{t-12}$$

Por tanto, la expresión algebraica de la temperatura mínima será:

$$X_t = 0,19848X_{t-1} + X_{t-12} - 0,19849X_{t-13} + Z_t - 0,88252Z_{t-12}$$

y la de la temperatura media será:

$$X_t = 0,19565X_{t-1} + X_{t-12} - 0,19565X_{t-13} + Z_t - 0,85557Z_{t-12}$$

A continuación, se va a analizar la estacionariedad de las precipitaciones. El correlograma simple tiende a 0 en la parte regular, pero no en la estacional, por lo que no es una serie estacionaria.

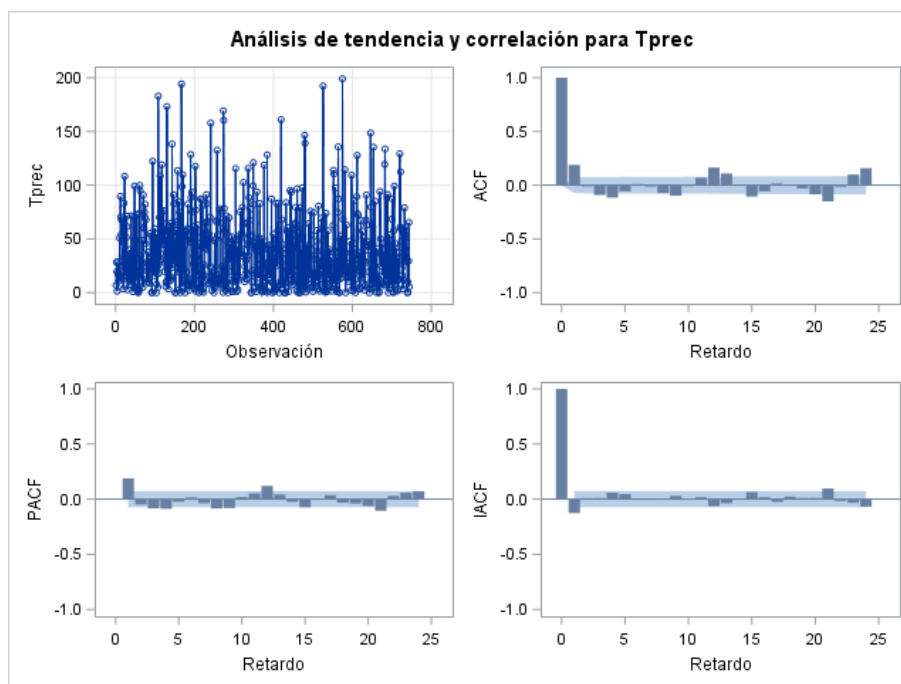


Figura 25. Representación gráfica de los correlogramas simple y parcial de las precipitaciones

Para transformarla en estacionaria, se va a diferenciar la parte estacional de la serie.

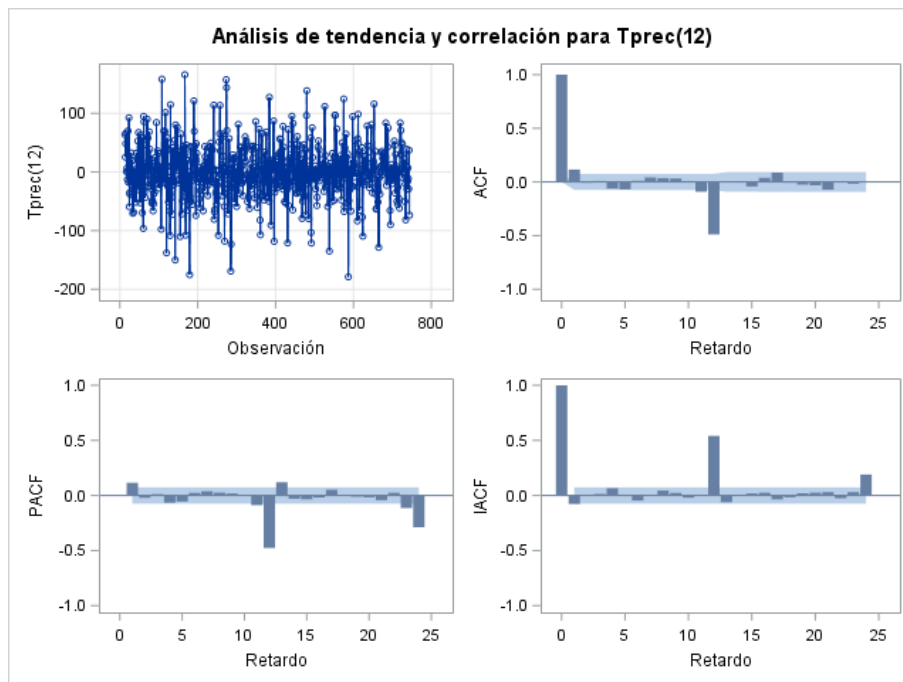


Figura 26. Funciones de autocorrelación simple y parcial de las precipitaciones con una diferencia en la parte estacional

La serie ya es estacionaria tanto en media como en varianza, por lo que se va a tratar de identificar el modelo ARIMA más adecuado. En la parte regular el único retardo levemente significativo tanto en las autocorrelaciones simples como en las parciales es el primero, lo cual puede indicar que la serie tenga un autorregresivo o una media móvil de orden 1. En la parte estacional, únicamente es significativo el primer retardo en las autocorrelaciones simples, pero en el caso de las parciales, los retardos decrecen tendiendo a 0, por lo que se podría tratar de una media móvil.

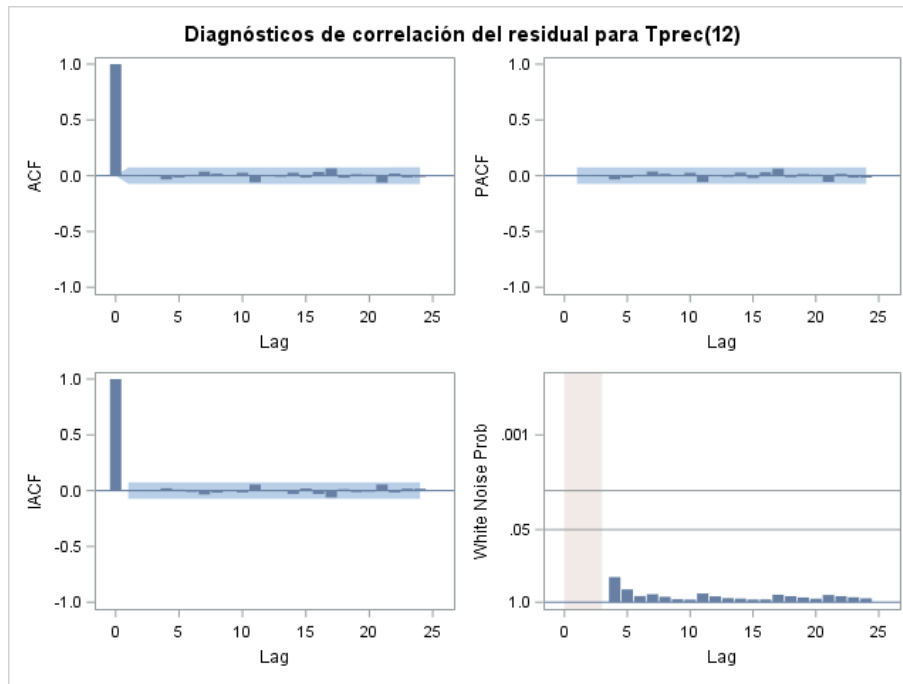


Figura 27. Funciones de autocorrelación simple y parcial de las precipitaciones con un autorregresivo y una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional

Todos los retardos de ambos correlogramas se anulan, pero el valor estimado tanto de la media como del autorregresivo y la media móvil de la parte regular no son significativos (tabla 16), es decir, se acepta la hipótesis nula de que esos valores son iguales a 0, por lo que deben ser eliminados del modelo.

Estimación por mínimos cuadrados condicional					
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t	Retardo
MU	0.09029	0.14713	0.61	0.5396	0
MA1,1	0.11521	0.44082	0.26	0.7939	1
MA2,1	0.90567	0.01592	56.89	<.0001	12
AR1,1	0.19805	0.43503	0.46	0.6491	1

Tabla 8. Estimaciones y significatividad de los parámetros del modelo ARIMA ajustado

En el caso del autorregresivo y la media móvil el problema podría encontrarse en que ambos están muy correlacionados (tabla 17), por lo que se va a eliminar únicamente la media móvil, ya que su valor estimado es mucho más próximo a 0.

Correlaciones de las estimaciones de parámetro				
Parámetro	MU	MA1,1	MA2,1	AR1,1
MU	1.000	0.004	0.111	0.004
MA1,1	0.004	1.000	0.056	0.996
MA2,1	0.111	0.056	1.000	0.056
AR1,1	0.004	0.996	0.056	1.000

Tabla 9. Correlaciones entre los parámetros del modelo ARIMA ajustado

Tras eliminar la constante y la media móvil de la parte regular, el autorregresivo pasa a ser significativo (tabla 18) y todos los retardos se anulan, por lo que el modelo explica toda la estructura de dependencia de la serie y esta sigue un proceso de ruido blanco.

Estimación por mínimos cuadrados condicional					
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t	Retardo
MA1,1	0.90410	0.01590	56.86	<.0001	12
AR1,1	0.08432	0.03694	2.28	0.0227	1

Tabla 10. Estimaciones y significatividad de los parámetros del modelo ARIMA ajustado

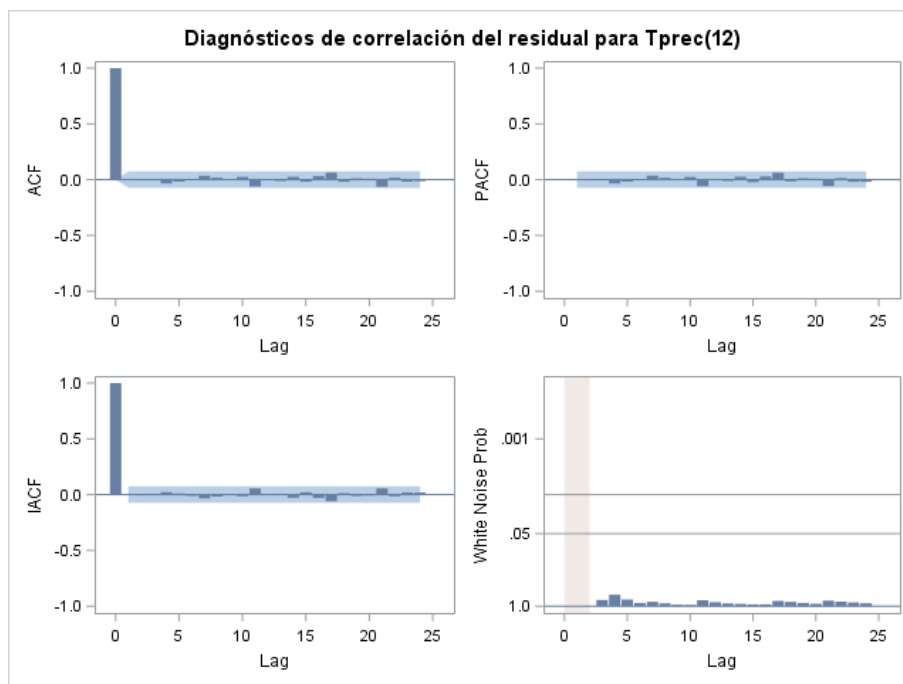


Figura 28. Funciones de autocorrelación simple y parcial de las precipitaciones con una media móvil en la parte regular y una diferencia y una media móvil en la parte estacional

Comprobación de autocorrelación de residuales										
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones						
6	1.11	4	0.8934	-0.001	0.009	0.010	-0.031	-0.015	0.010	
12	5.56	10	0.8508	0.037	0.020	0.002	0.027	-0.059	-0.002	
18	10.73	16	0.8258	-0.009	0.028	-0.016	0.034	0.066	-0.018	
24	14.44	22	0.8853	0.014	0.011	-0.061	0.020	-0.016	-0.011	
30	27.67	28	0.4820	-0.103	0.051	0.050	0.028	-0.026	-0.018	
36	35.44	34	0.4002	0.038	0.051	0.020	-0.002	0.062	-0.043	
42	42.01	40	0.3839	0.011	0.075	0.001	0.051	0.011	0.002	
48	51.07	46	0.2812	0.063	-0.026	0.034	0.060	0.047	-0.007	

Tabla 11. Contrastes de aleatoriedad de los residuos del modelo ARIMA ajustado

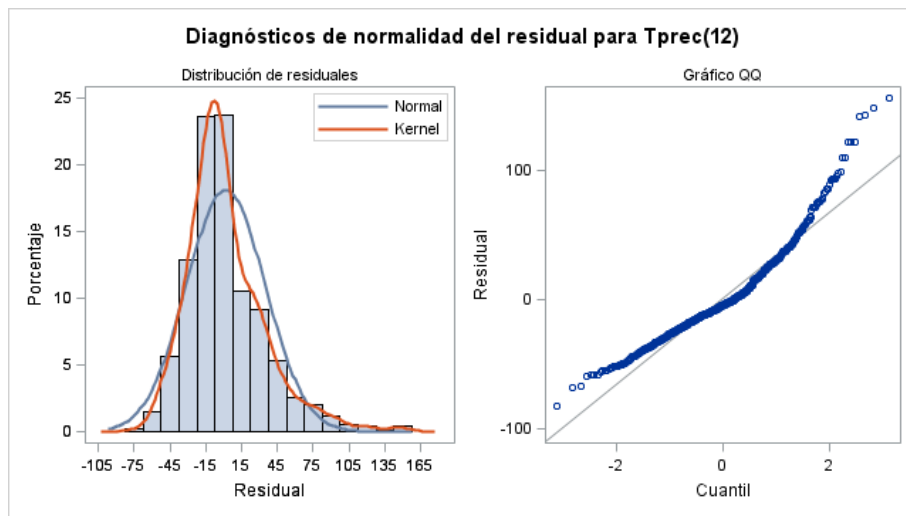


Figura 29. Diagramas de normalidad de los residuos de las precipitaciones

Por tanto, el modelo ARIMA más adecuado para realizar la predicción de la serie de precipitaciones totales va a ser un modelo diferenciado en la parte estacional con un autorregresivo en la parte regular y una media móvil en la parte estacional, el cual coincide con el modelo de las temperaturas: ARIMA (1,0,0) (0,1,1)₁₂.

Los parámetros estimados del modelo ajustado son los siguientes:

Factores autoregresivos	
Factor 1:	1 - 0.08432 B**(1)
Factores de la media móvil	
Factor 1:	1 - 0.9041 B**(12)

Tabla 12. Parámetros estimados del modelo ajustado para las precipitaciones

Con ellos se obtiene la expresión algebraica de este modelo ajustado:

$$(1 - \phi_1 B)(1 - B^{12})X_t = (1 - \theta_1 B^{12})Z_t$$

$$(1 - 0,0832B)(1 - B^{12})X_t = (1 - 0,9041B^{12})Z_t$$

$$(1 - 0,0832B - B^{12} + 0,0832B^{13})X_t = (1 - 0,9041B^{12})Z_t$$

$$X_t - 0,0832X_{t-1} - X_{t-12} + 0,0832X_{t-13} = Z_t - 0,9041Z_{t-12}$$

$$X_t = 0,0832X_{t-1} + X_{t-12} - 0,0832X_{t-13} + Z_t - 0,9041Z_{t-12}$$

Se ha comprobado, además, que los modelos más adecuados para las estaciones de Barcelona y Huelva seleccionadas son los mismos que los obtenidos para la estación de Madrid, por lo que se presupone que se ajustarán correctamente al resto de estaciones meteorológicas.

5.2.3. Cálculo y representación de las predicciones

A partir de los modelos ARIMA ajustados, se van a realizar predicciones de las temperaturas máximas, mínimas y medias y de las precipitaciones totales de los 6 años siguientes al último periodo observado, es decir, desde enero de 2012 hasta diciembre de 2017. A continuación, se representan tanto los residuales como las predicciones realizadas.

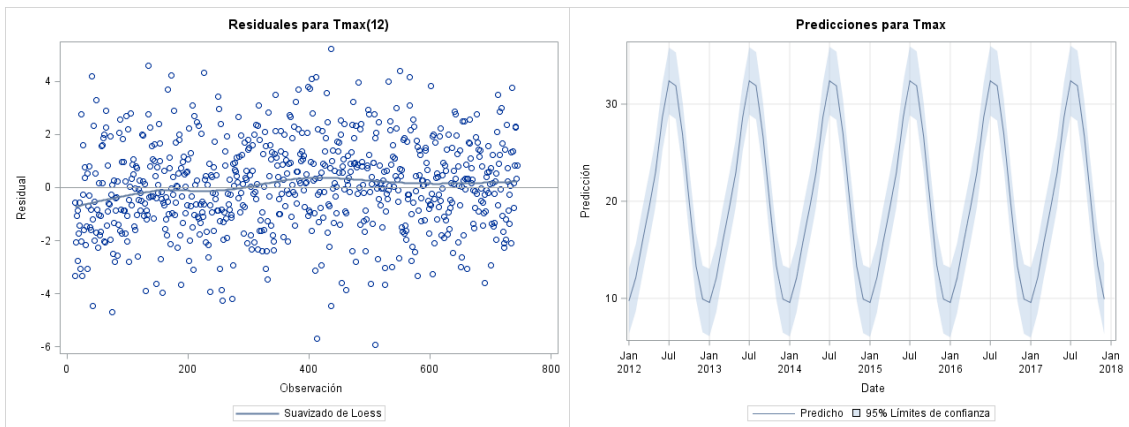


Figura 30. Residuales y predicciones de la temperatura máxima

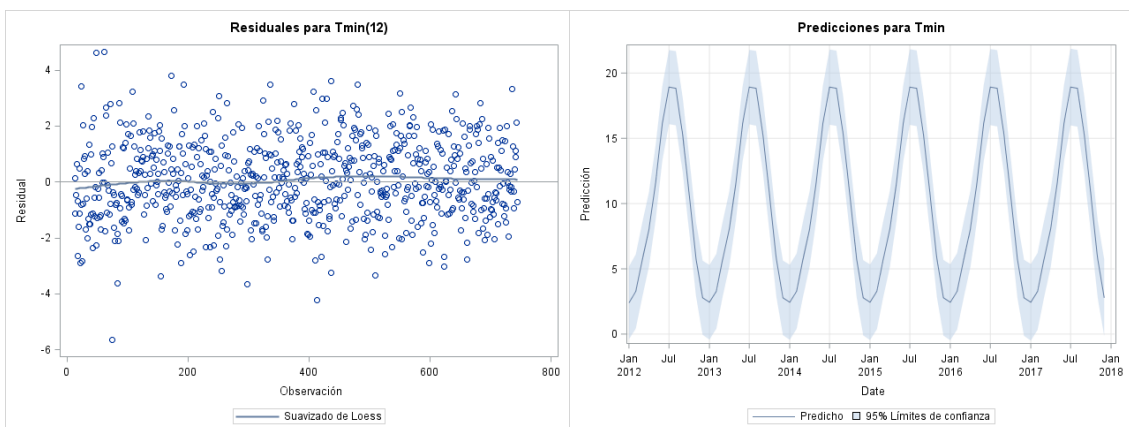


Figura 31. Residuales y predicciones de la temperatura mínima

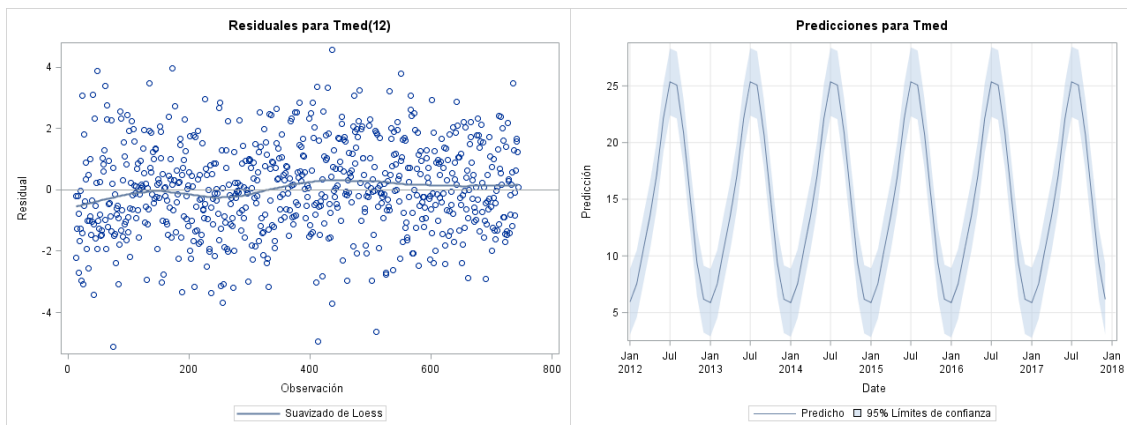


Figura 32. Residuales y predicciones de la temperatura media

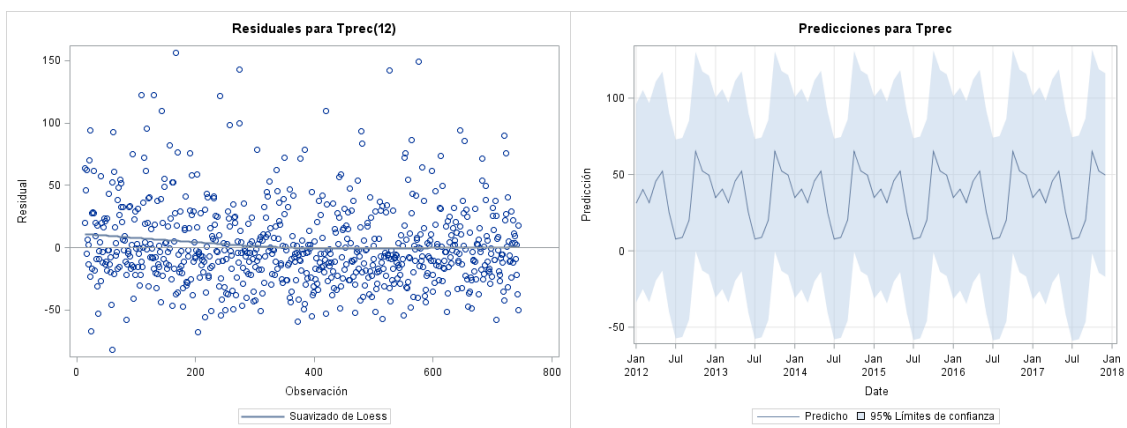


Figura 33. Residuales y predicciones de las precipitaciones totales

Los residuales de las temperaturas se sitúan en torno a 0 y sus intervalos de confianza son bastante estrechos, lo que indica que el modelo ajustado permite realizar unas buenas predicciones. En cambio, en el caso de las precipitaciones, los residuales tienen bastantes valores alejados de 0 y las bandas de confianza son muy amplias, ya que el modelo no recoge la gran cantidad de datos atípicos que existen, por lo que sus predicciones podrían no ser muy acertadas.

A partir de los datos mensuales reales del periodo 2014-2017 se va a comprobar la fiabilidad de las predicciones realizadas para todas las estaciones meteorológicas en conjunto.

El coeficiente de correlación de Pearson va a permitir comparar los datos estimados con los reales. En cuanto a las temperaturas, este coeficiente es muy elevado, 0,972 en las máximas, 0,971 en las mínimas y 0,979 en las medias, por lo que se podría decir que los modelos ajustados para realizar la predicción de estas tres variables funcionan correctamente. Además, los gráficos de dispersión entre cada par de variables muestran una clara relación lineal.

En cambio, el coeficiente de correlación de las precipitaciones, 0,613, indica que el modelo ajustado para realizar la predicción de las mismas no funciona bien del todo, ya que existen demasiados valores atípicos que no son recogidos por el modelo, lo cual distorsiona la predicción. El gráfico de dispersión entre la variable real y la estimada confirma este resultado.

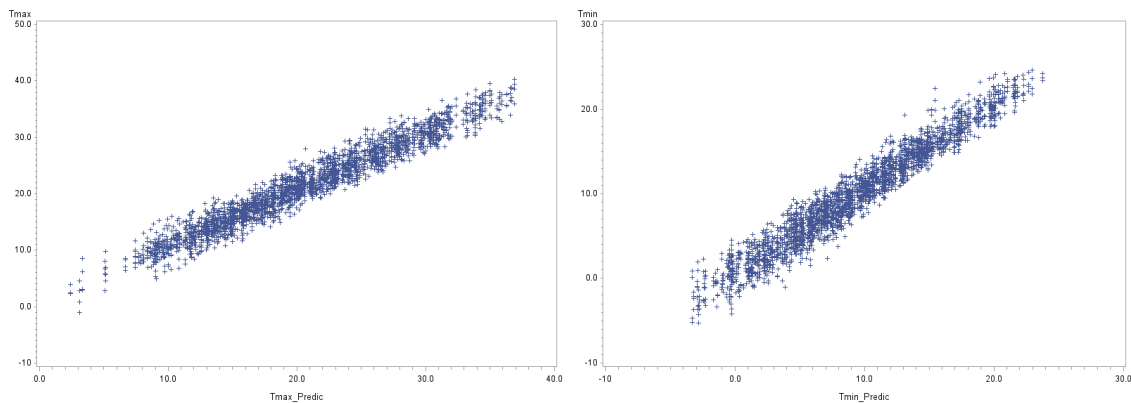


Figura 34. Gráfico de dispersión entre la temperatura máxima estimada y la real (izquierda) y entre la temperatura mínima estimada y la real (derecha)

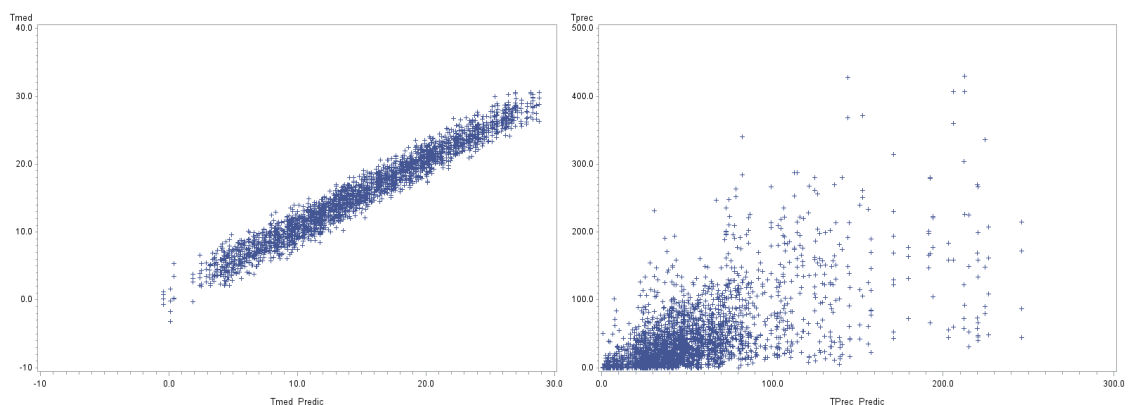


Figura 35. Gráfico de dispersión entre la temperatura media estimada y la real (izquierda) y entre las precipitaciones estimadas y las reales (derecha)

5.3. Análisis de la tendencia

5.3.1. Cálculo y representación de la tendencia en mapas

Analizar la tendencia que han seguido las diferentes variables climáticas va a permitir conocer con mayor precisión si sus valores han ido creciendo, disminuyendo o se han mantenido constantes a lo largo de los años. Este análisis se va a realizar a nivel de estación y mes sobre las 56 series seleccionadas previamente, comunes a los dos periodos de estudio, por lo que se va a obtener una única tendencia por cada mes y estación. Los resultados se presentan en forma de mapas, los cuales representan la tendencia que tienen las temperaturas en cada provincia y en cada uno de los meses del año.

En primer lugar, se calcula la correlación de primer orden para cada serie y mes. En aquellos casos en los que exista esta correlación, se preblanquea la serie para eliminarla, de modo que el valor original de las temperaturas/precipitaciones es sustituido por el valor preblanqueado.

Una vez eliminadas las correlaciones de todas las series, se aplica el test de Mann-Kendall para determinar las tendencias significativas. Este contraste se realiza con los datos diferenciados para eliminar la estacionalidad de las series.

Las precipitaciones, en general, no tienen tendencia, ya que tan solo 22 de las 672 series (56 estaciones por 12 meses) la tienen, por lo que no tiene sentido continuar con su análisis. En cambio, las temperaturas muestran una tendencia significativa en la mayoría de las series (453 en las temperaturas máximas, 351 en las mínimas y 405 en las medias), por lo que el siguiente paso va a ser calcular el valor de la pendiente de cada tendencia a partir del test de Theil-Sen. En aquellas series que carezcan de tendencia, el valor asociado a la pendiente será 0. El hecho de que los años partan de 1950 afecta a los valores de la tendencia, por lo que, en su lugar, se va a partir de 0 para que los datos tengan sentido.

Los valores calculados de las tendencias se muestran a través de mapas. Cada uno de ellos representa la tendencia de la temperatura máxima, mínima o media en cada provincia de la España peninsular en un mes del año. Cada una de estas provincias es representada por una o varias de las 56 estaciones meteorológicas analizadas. Estas estaciones abarcan 42 de las 47 provincias, ya que ninguna de ellas pertenece a Álava, Guipúzcoa, Lleida, Málaga y Palencia, por lo que sus tendencias no han podido ser calculadas. También hay casos en los que se ha utilizado más de una estación por provincia, por lo que la tendencia ha sido calculada como el promedio de los valores de las diferentes estaciones pertenecientes a esas provincias, las cuales son las siguientes:

- A Coruña: 3 estaciones (1387, 1387E, 1428).
- Alicante: 2 estaciones (8019, 8025).
- Asturias: 2 estaciones (1212E, 1249I).
- Cantabria: 2 estaciones (1109, 1111).
- Cádiz: 3 estaciones (5960, 5973, 6001).
- Madrid: 4 estaciones (2462, 3129, 3191E, 3195).
- Pontevedra: 2 estaciones (1484C, 1495).
- Salamanca: 2 estaciones (2867, 2870).
- Sevilla: 2 estaciones (5783, 5796).
- Teruel: 2 estaciones (8368U, 9381I).

El color blanco representa las provincias que no tienen tendencia y el color rojo representa tendencias positivas; cuanto mayor es la pendiente de la tendencia, más intenso es este color, alcanzando la máxima intensidad en valores situados entre 0,040916 y 0,065975. Se han ajustado un total de 11 categorías de colores comunes a todos los meses y tipos de temperaturas (máxima, mínima y media) para poder realizar un análisis conjunto, de manera que un determinado valor siempre tenga el mismo color en múltiples mapas.

A modo de ejemplo, se presenta el mapa de las tendencias de la temperatura máxima en el mes de enero, en el que se pueden diferenciar los valores de la pendiente asociados a cada tono.

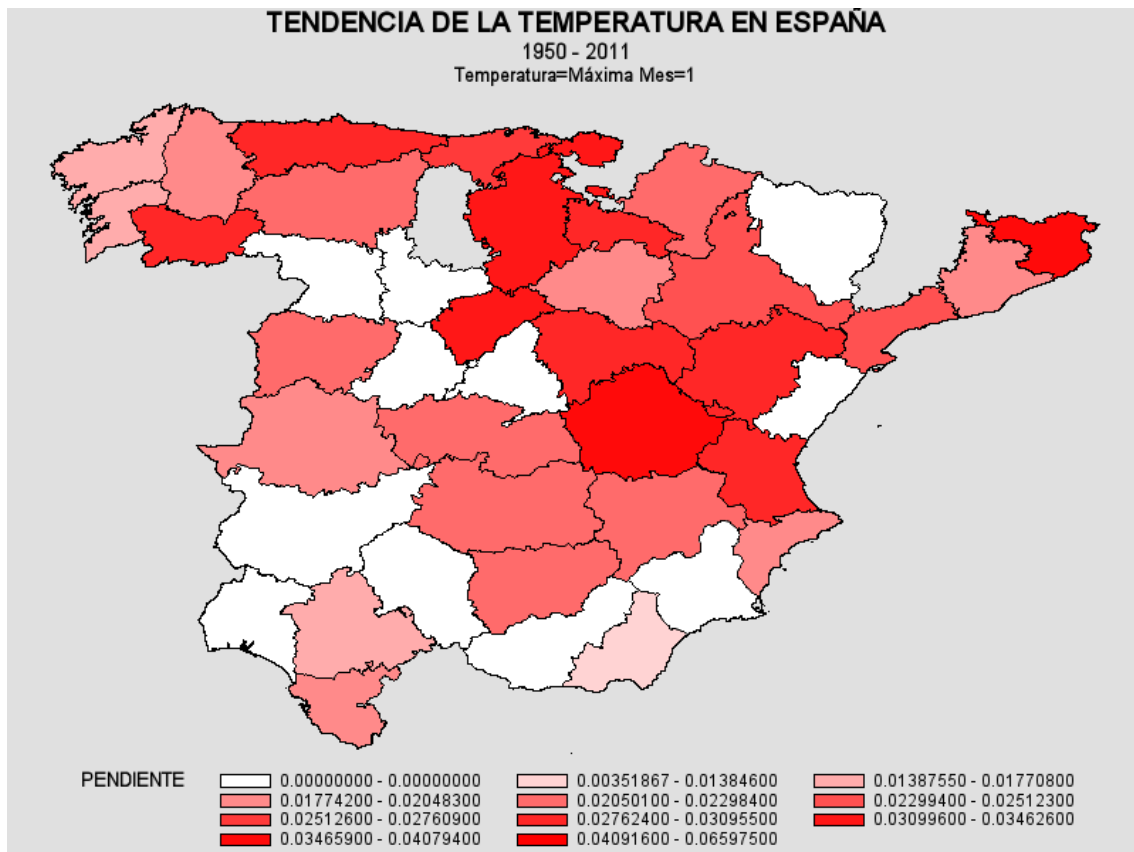


Figura 36. Tendencia de la temperatura máxima en la España peninsular en el mes de enero desde 1950 hasta 2011

A continuación, se presentan los 12 mapas de tendencias de la temperatura máxima, de la mínima y de la media. Como se puede observar, ninguna de las tendencias es negativa, lo que indica que, desde el año 1950 hasta el año 2011, las temperaturas de la España peninsular se han mantenido constantes o han tendido a crecer en mayor o menor medida. Esta afirmación corrobora la hipótesis planteada al inicio del estudio de que “la temperatura en España tiene una tendencia creciente”.

En cuanto a las temperaturas máximas, los meses en los que más han tendido a crecer las mismas son febrero, marzo, julio y agosto, mientras que septiembre y noviembre han sido los más constantes. En relación a las temperaturas mínimas, se observa un claro aumento gradual de las tendencias desde el mes de noviembre hasta el mes de junio, mientras que el resto de los meses no siguen un patrón claro. Por último, las tendencias de las temperaturas medias son muy similares a las de las máximas; tienden a crecer en mayor medida en los meses de junio y agosto, pero se mantienen estables en septiembre, noviembre y diciembre.

En general, se puede apreciar que tanto las temperaturas máximas como las medias tienen una tendencia creciente mucho mayor que las temperaturas mínimas. El mes de noviembre es aquel en el que más estables se han mantenido las temperaturas a lo largo de los años, mientras que en junio se han producido los mayores crecimientos de la tendencia, es decir, cada año que pasa, los meses de junio se vuelven más calurosos, pero los meses de noviembre suelen tener unas temperaturas similares.

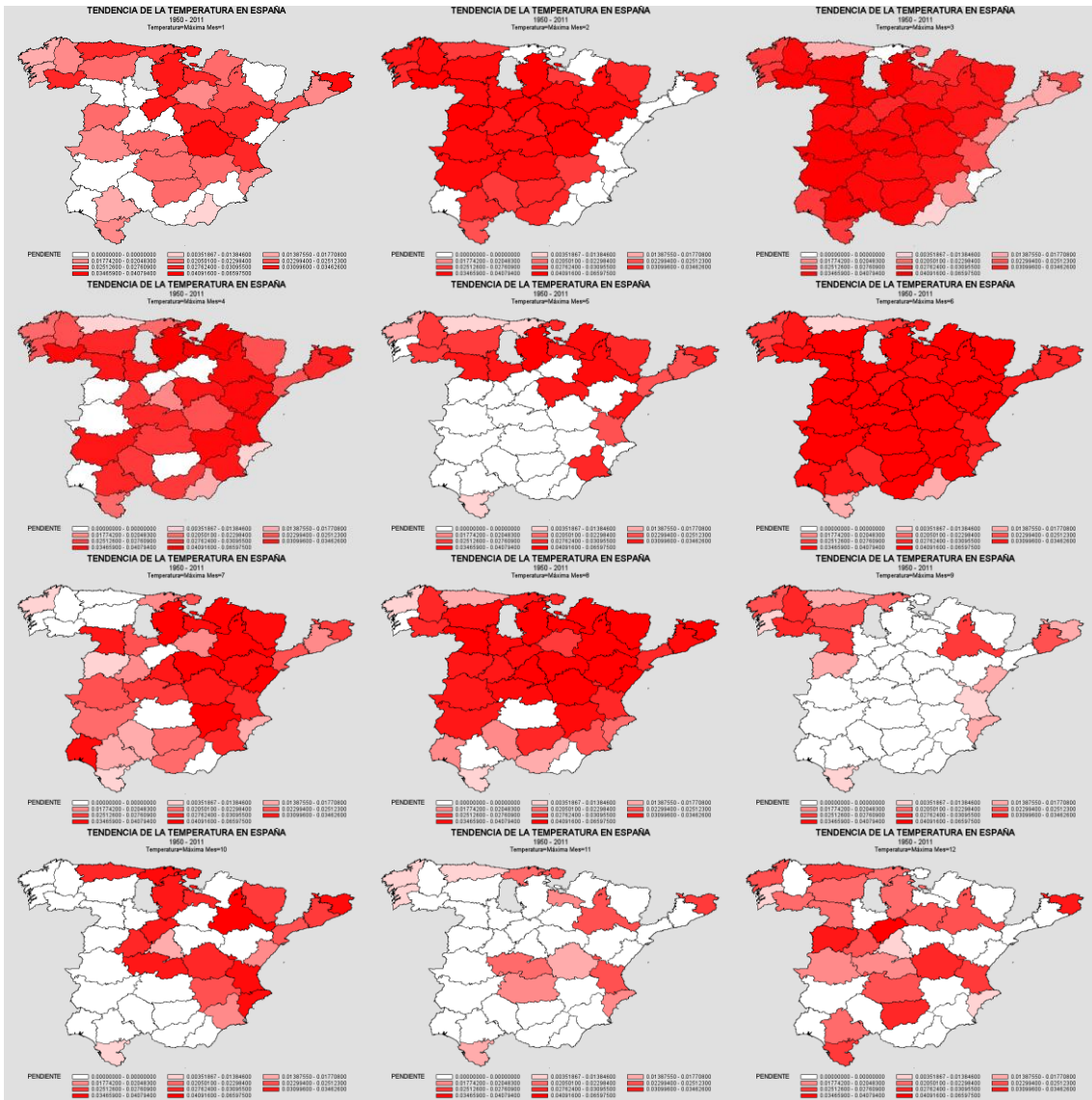


Figura 37. Tendencia de la temperatura máxima en la España peninsular en cada uno de los meses del año (ordenados de enero a diciembre) desde 1950 hasta 2011

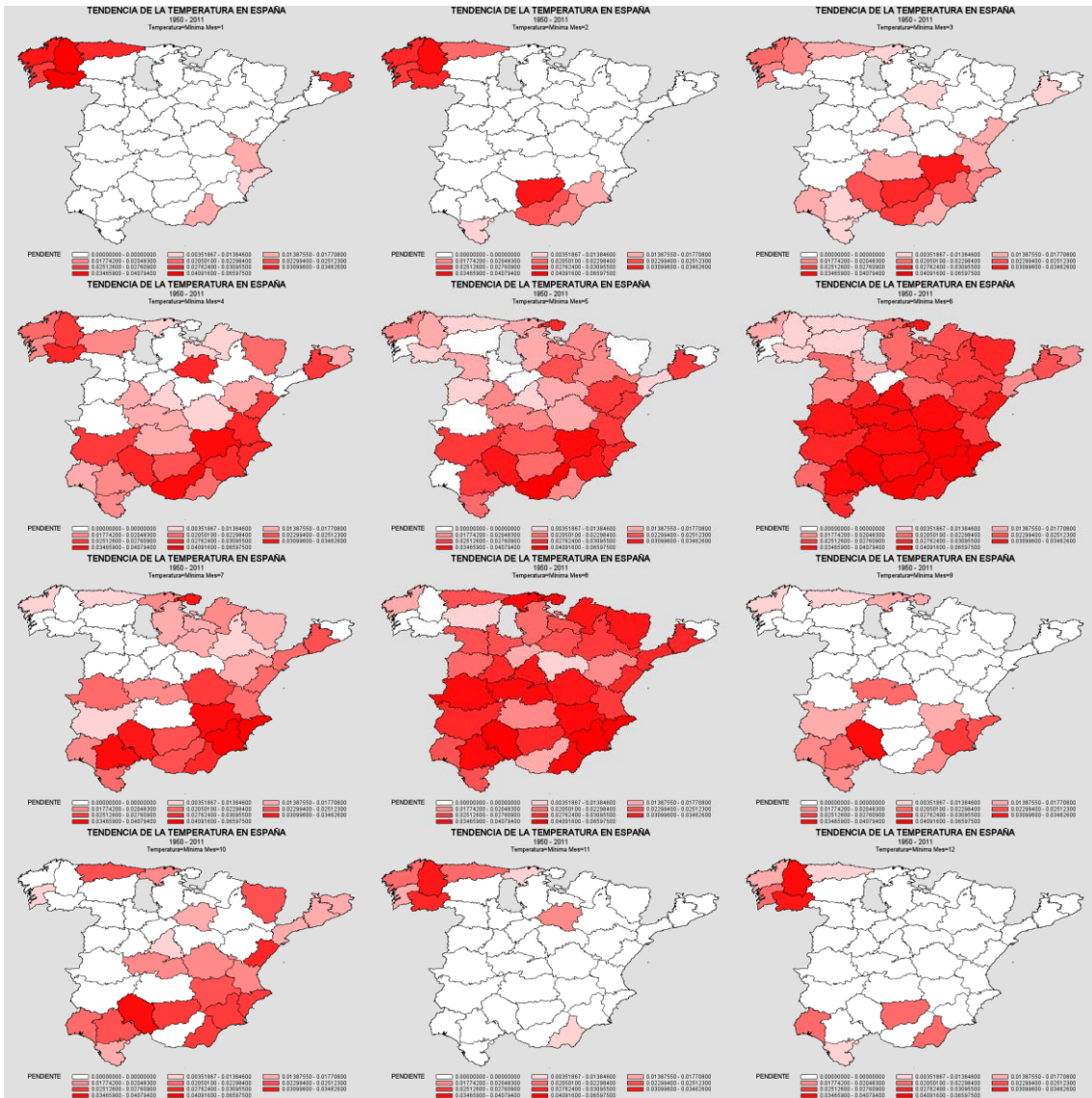


Figura 38. Tendencia de la temperatura mínima en la España peninsular en cada uno de los meses del año (ordenados de enero a diciembre) desde 1950 hasta 2011

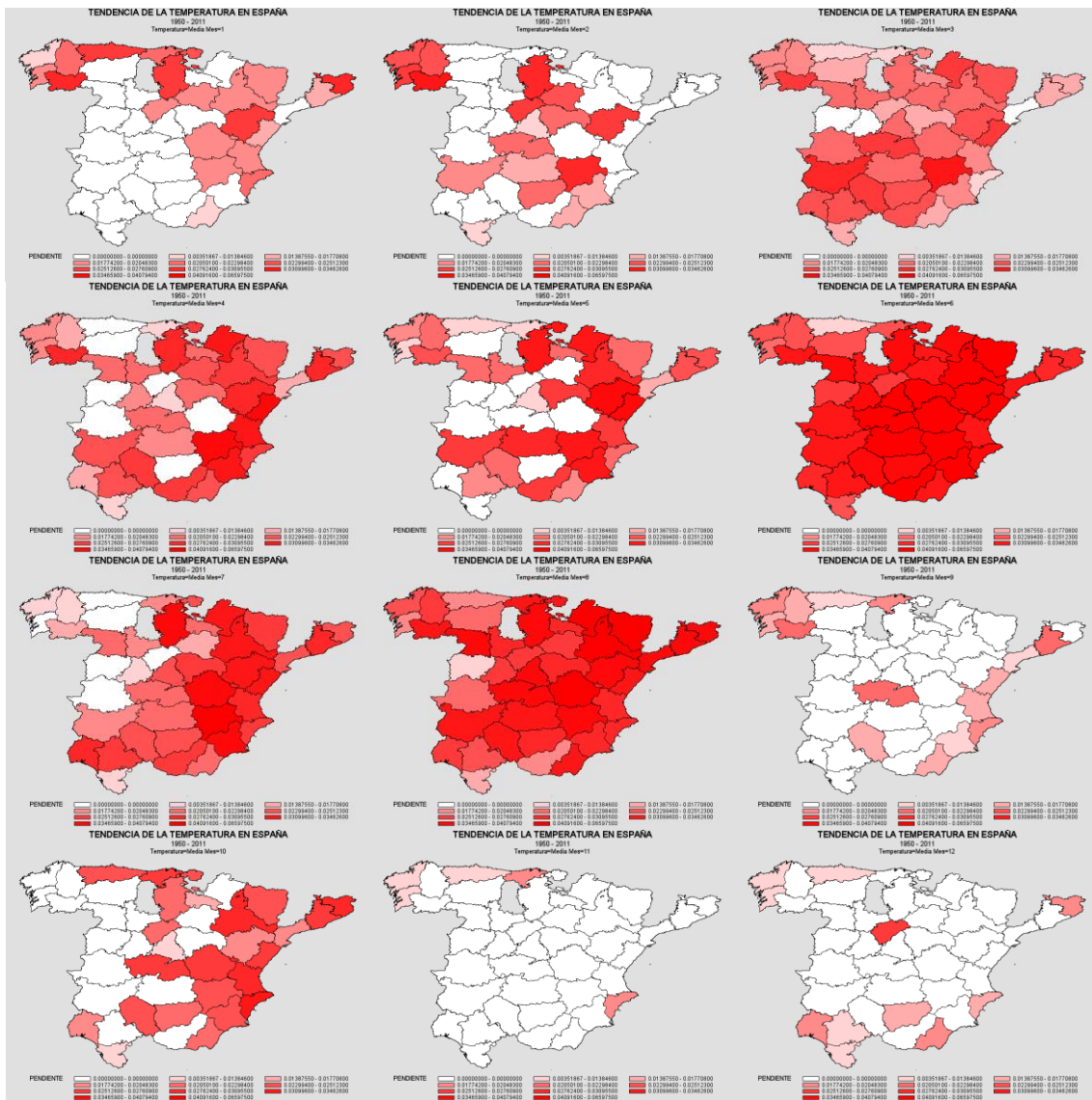


Figura 39. Tendencia de la temperatura media en la España peninsular en cada uno de los meses del año (ordenados de enero a diciembre) desde 1950 hasta 2011

5.3.2. Regresión lineal

Por otro lado, se han construido una serie de modelos de regresión lineal en los que la tendencia de la temperatura máxima, mínima o media en cada uno de los meses del año depende de la altitud, de las coordenadas cartesianas (x, y) y de la distancia al mar de las estaciones meteorológicas, con el fin de detectar la influencia que tiene cada una de estas variables sobre la tendencia. La latitud y la longitud han sido transformadas en estas coordenadas cartesianas. Además, se ha decidido no realizar una selección de variables.

A continuación, se presentan los estimadores de los parámetros de cada una de las regresiones, así como la raíz del error cuadrático medio que cometen.

Tendencia	Mes	\sqrt{ECM}	T. indep.	Altitud	Coordenada X	Coordenada Y	Distancia mar
Tmax	1	0.010512	-0.05137	0.0000029	0.0000000094	0.0000000149	-0.00002794
	2	0.010000	0.02147	0.0000175	-0.0000000301	0.0000000012	0.00004195
	3	0.009530	0.06628	0.0000046	-0.0000000170	-0.0000000084	0.00006234
	4	0.013211	-0.02251	0.0000004	0.0000000151	0.0000000091	-0.00000756
	5	0.013050	-0.11596	0.0000088	0.0000000206	0.0000000270	-0.00004183
	6	0.010080	0.01452	0.0000101	0.0000000193	0.0000000010	0.00007574
	7	0.013163	0.00132	0.0000073	0.0000000372	-0.0000000003	0.00002484
	8	0.010736	-0.11115	0.0000025	0.0000000375	0.0000000259	0.00006608
	9	0.010180	-0.08042	-0.0000085	-0.0000000064	0.0000000214	-0.00000271
	10	0.012613	-0.08315	-0.0000025	0.0000000409	0.0000000182	-0.00000060
	11	0.009339	-0.01702	-0.0000093	0.0000000092	0.0000000052	0.00000445
	12	0.012656	0.01908	0.0000094	-0.0000000162	0.0000000001	-0.00001607
Tmin	1	0.010226	-0.07621	-0.0000006	-0.0000000217	0.0000000218	-0.00004919
	2	0.008884	0.01694	0.0000156	-0.0000000277	0.0000000009	-0.00008074
	3	0.008830	0.08280	0.0000111	-0.0000000104	-0.0000000147	-0.00006999
	4	0.009987	0.10144	-0.0000013	0.0000000010	-0.0000000182	-0.00003387
	5	0.010659	0.09773	-0.0000014	0.0000000120	-0.0000000185	-0.00001864
	6	0.008322	0.14915	-0.0000025	0.0000000193	-0.0000000294	0.00002373
	7	0.009523	0.12785	-0.0000116	0.0000000222	-0.0000000260	0.00000565
	8	0.010212	0.07915	-0.0000122	0.0000000166	-0.0000000137	0.00005528
	9	0.008050	0.09089	-0.0000162	-0.0000000016	-0.0000000177	0.00003395
	10	0.009442	0.08046	-0.0000108	0.0000000156	-0.0000000157	-0.00000875
	11	0.006400	-0.06038	0.0000021	-0.0000000195	0.0000000171	-0.00003738
	12	0.008051	0.00670	0.0000030	-0.0000000228	0.0000000025	-0.00004223
Tmed	1	0.008384	-0.08382	0.0000039	0.0000000095	0.0000000209	-0.00005300
	2	0.010806	0.01729	0.0000214	-0.0000000170	-0.0000000009	-0.00004733
	3	0.008519	0.05035	0.0000055	-0.0000000023	-0.0000000073	-0.00000362
	4	0.010841	0.02742	0.0000010	0.0000000182	-0.0000000031	-0.00002657
	5	0.011803	-0.02910	-0.0000049	0.0000000244	0.0000000085	-0.00001174
	6	0.007747	0.08554	0.0000061	0.0000000218	-0.0000000148	0.00003099
	7	0.009342	0.05783	0.0000032	0.0000000320	-0.0000000116	-0.00000311
	8	0.006875	-0.01763	-0.0000019	0.0000000230	0.0000000079	0.00002887
	9	0.007712	-0.04740	-0.0000145	0.0000000012	0.0000000131	0.00001515
	10	0.009162	-0.01588	-0.0000061	0.0000000362	0.0000000040	-0.00001717
	11	0.005610	-0.02191	-0.0000023	-0.0000000002	0.0000000060	-0.00001392
	12	0.007869	0.05731	0.0000025	-0.0000000077	-0.0000000103	-0.00003458

Tabla 13. Estimadores de los parámetros de la regresión lineal

Cuando un parámetro tiene una estimación positiva/negativa, un aumento del mismo incrementa/disminuye la tendencia, siempre que el resto de variables se mantengan constantes.

Se utiliza el modelo de la tendencia de la temperatura máxima en enero como ejemplo para representar la expresión de la regresión lineal:

$$\begin{aligned} & \text{Tendencia de la temperatura máxima en enero} \\ & = -0,05137 + 0,0000029 * \text{Altitud} + 0,000000094 * X \\ & + 0,0000000149 * Y - 0,00002794 * \text{Distancia al mar} \end{aligned}$$

Cuanto más alta esté situada una estación, mayor será la tendencia de la temperatura máxima en enero. Lo mismo ocurre con las coordenadas cartesianas, ya que cuanto más a la derecha (X) o más arriba (Y) esté localizada una estación en el mapa, mayor será la tendencia. En cambio, la relación entre la tendencia y la distancia al mar es negativa, ya que cuanto más lejos esté una estación del mar, menor será esta tendencia.

En el caso de las temperaturas máximas, a mayor altitud, mayor es la tendencia en todos los meses salvo en octubre, septiembre y noviembre, en los que es menor. Cuanto más a la derecha del mapa está localizada una estación, mayor es la tendencia en enero, abril, mayo, junio, julio, agosto, octubre y noviembre, pero menor lo es en el resto de meses, cuanto más arriba del mapa está localizada una estación, mayor es la tendencia en todos los meses salvo en marzo y julio, y cuanto más alejada está una estación del mar, mayor es la tendencia en febrero, marzo, junio, julio, agosto y noviembre, pero menor lo es en el resto de meses.

En cuanto a las temperaturas mínimas, a mayor altitud, menor es la tendencia en todos los meses salvo en febrero, marzo, noviembre y diciembre, en los que es mayor. Cuanto más a la derecha del mapa está localizada una estación, mayor es la tendencia en abril, mayo, junio, julio, agosto y octubre, pero menor lo es en el resto de meses, cuanto más arriba del mapa está localizada una estación, menor es la tendencia en todos los meses salvo en enero, febrero, noviembre y diciembre, y cuanto más alejada está una estación del mar, menor es la tendencia en todos los meses salvo en junio, julio, agosto y septiembre.

Por último, en el caso de las temperaturas medias, a mayor altitud, menor es la tendencia en mayo, agosto, septiembre, octubre y noviembre, pero mayor lo es en el resto de meses. Cuanto más a la derecha del mapa está localizada una estación, mayor es la tendencia en todos los meses salvo en febrero, marzo, noviembre y diciembre, en los que es menor. Cuanto más arriba del mapa está localizada una estación, mayor es la tendencia en enero, mayo, agosto, septiembre, octubre y noviembre, pero menor lo es en el resto de meses, y cuanto más alejada está una estación del mar, menor es la tendencia en todos los meses salvo en junio, agosto y septiembre.

En general, la variable que más influencia tiene sobre la tendencia de las temperaturas es la distancia al mar, ya que sus estimadores son mayores que los del resto de variables, seguida de la altitud, y las variables menos influyentes son las coordenadas cartesianas.

6. CONCLUSIONES

La consecución de los objetivos propuestos al inicio del estudio ha sido satisfactoria. Se ha conseguido conocer el comportamiento climático de la España peninsular desde el año 1950 hasta el año 2011 a partir del análisis de la evolución de las temperaturas y de las precipitaciones, lo que ha permitido conocer sus valores futuros.

Para ello, siguiendo los objetivos específicos establecidos, se han generado dos conjuntos de datos (1950-2011 y 2014-2017) debidamente homogeneizados, en los que las alteraciones producidas en los datos tan solo se deben a factores climáticos. También se han completado los valores ausentes y se han corregido los datos atípicos considerados anómalos a partir de diversas series de referencia construidas en función de los valores de las series más correlacionadas con la serie objetivo.

El conjunto de datos del periodo 1950-2011 ha sido analizado para ajustar un modelo que se adapte a los datos, el cual permite conocer el comportamiento que seguirán en un futuro las temperaturas máximas, mínimas y medias, pero, en el caso de las precipitaciones, la presencia de tantos valores atípicos ha impedido la generación de un buen modelo predictivo. Por tanto, la hipótesis planteada al inicio del estudio de que “es posible predecir la temperatura mínima, máxima y media y las precipitaciones de un lugar de la España peninsular en un determinado periodo de tiempo” ha sido confirmada únicamente en lo que respecta a las temperaturas. Estas tienen un comportamiento estacional regular muy marcado cada 12 meses.

Tanto este análisis de la evolución de las series como el análisis de las tendencias han permitido corroborar la hipótesis planteada de que “la temperatura en España tiene una tendencia creciente”, ya que ninguna de las tendencias es negativa, sino que las temperaturas de la España peninsular han tenido una tendencia creciente en la mayoría de las provincias y de los meses del año, siendo junio el mes con las tendencias más elevadas y, noviembre, el mes con las tendencias más constantes. En cambio, la hipótesis de que “las precipitaciones han ido disminuyendo a lo largo de los años en España” no ha sido confirmada, ya que se ha concluido que las precipitaciones no tienen ningún tipo de tendencia a nivel general, ya que su comportamiento a lo largo del periodo de estudio ha sido constante.

En cuanto a la hipótesis de que “la localización de las estaciones meteorológicas influye sobre la tendencia que han seguido las temperaturas y las precipitaciones en la España peninsular en los últimos años”, la tendencia de las temperaturas depende principalmente de la distancia a la que se encuentran las estaciones meteorológicas del mar y de la altitud a la que están situadas.

En definitiva, en la España peninsular las temperaturas han tenido una tendencia creciente durante las últimas décadas, dando lugar al calentamiento global del que tanto se habla.

En cuanto a líneas futuras, se podría plantear un modelo de regresión en el que se relacionen las temperaturas con unos índices que recogen las diferencias de presión existentes en el planeta, cuya probable influencia revelaría una de las posibles causas

de la creciente tendencia detectada en las temperaturas. También podrían analizarse otro tipo de variables climáticas, como las horas de sol, la velocidad del viento o las rachas de viento.

7. BIBLIOGRAFÍA

- [1] Alonso Revenga, J.M. 2018. Análisis Descriptivo de una serie temporal. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid: Material de la asignatura Complementos de Formación en Técnicas de Minería de Datos.
- [2] Alonso Revenga, J.M. 2018. Identificación, estimación, diagnóstico y predicción de modelos ARIMA. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid: Material de la asignatura Complementos de Formación en Técnicas de Minería de Datos.
- [3] Alonso Revenga, J.M. 2018. Modelos de series temporales estacionarias. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid: Material de la asignatura Complementos de Formación en Técnicas de Minería de Datos.
- [4] Calviño, A. 2018. Regresión lineal. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid: Material de la asignatura Técnicas y Metodología de la Minería de Datos (SEMMA).
- [5] CARTAYA, S. 2016. Métodos de ajuste y homogenización de datos climáticos para determinar índice de humedad de Lang en la provincia de Manab. *La Técnica*, 16, 94-106. ISSN: 1390-6895.
- [6] DE CASTRO, M., MARTÍN-VIDE, J. & ALONSO, S. 2005. El clima de España: pasado, presente y escenarios de clima para el siglo XXI. España: Ministerio de Medio Ambiente. ISBN 84-8320-303-0.
- [7] DE LUQUE SÖLLHEIM, A.L. 2011. Cualificación y homogenización de las series climáticas mensuales de precipitación de Canarias; Memoria Metodológica. Agencia Canaria de Desarrollo Sostenible y Cambio Climático.
- [8] ESTEBAN-PARRA, M.J. 1995. Contribución al estudio de la evolución del clima de España en el periodo instrumental. Tesis doctoral, Universidad de Granada.
- [9] FONT TULLOT, I. 2000. Climatología de España y Portugal. 2ª ed. Universidad de Salamanca. ISBN 84-7800-944-2.
- [10] GUIJARRO, J.A. 2018. Homogeneización de series climáticas con Climatol. Agencia Estatal de Meteorología (AEMET), D.T. en Islas Baleares, España. Versión 3.1.1.
- [11] GUTIÉRREZ TEIRA, A. & PICATOSTE RUGGERONI, J.R. 2012. *Evidencias del cambio climático y sus efectos en España*. Oficina Española de Cambio Climático, Ministerio de Medio Ambiente y Medio Rural y Marino (MAGRAMA).
- [12] LÓPEZ, P.M. [et al.] 2015. Análisis de tendencia en series autocorrelacionadas. ResearchGate.
- [13] RÍOS CORNEJO, D. 2015. Variaciones climáticas y su posible influencia en la vegetación potencial natural de la España peninsular. Ángel Penas Merino (dir.) & Sara del Río González (dir.). Tesis doctoral, Universidad de León.

- [14] SHENG, Y. 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes*, 16(9), 1807-1829.
- [15] STEPÁNEK, P. 2003. Homogeneización de las series de temperatura del aire en la República Checa en el periodo instrumental. *Geographica*, 43, 5-24. ISSN 0210-8380.

8. ANEXOS

A. Tablas

Estaciones meteorológicas del periodo 1950-2011

ID	Nombre	Provincia	Altitud	Latitud	Longitud
0016A	Reus Aeropuerto	Tarragona	71	410859N	011044E
0076	Barcelona Aeropuerto	Barcelona	4	411734N	020412E
0200E	Barcelona, Fabra	Barcelona	412	412506N	020727E
0367	Girona Aeropuerto	Girona	143	415442N	024548E
1014	Hondarribia, Malkarroa	Gipuzkoa	4	432125N	014732W
1024E	San Sebastián, Igueldo	Gipuzkoa	251	431823N	020228W
1082	Bilbao Aeropuerto	Bizkaia	42	431753N	025421W
1109	Santander Aeropuerto	Cantabria	5	432545N	034953W
1110	Santander, Ciudad	Cantabria	64	432754N	034908W
1111	Santander	Cantabria	52	432928N	034802W
1208H	Gijón, Puerto	Asturias	5	433336N	054203W
1212E	Asturias Aeropuerto	Asturias	127	433401N	060239W
1249I	Oviedo	Asturias	336	432112N	055227W
1387	A Coruña	A Coruña	58	432157N	082517W
1387E	A Coruña Aeropuerto	A Coruña	98	431825N	082219W
1428	S. de Compostela Aeropuerto	A Coruña	370	425317N	082438W
1484C	Pontevedra	Pontevedra	108	422618N	083657W
1495	Vigo Aeropuerto	Pontevedra	261	421419N	083726W
1505	Lugo Aeropuerto	Lugo	445	430641N	072727W
1549	Ponferrada	León	534	423350N	063600W
1690A	Ourense	Ourense	143	421931N	075135W
2030	Soria	Soria	1082	414630N	022859W
2331	Burgos Aeropuerto	Burgos	891	422122N	033717W
2401	Palencia, Observatorio	Palencia	750	420018N	043145W
2422	Valladolid	Valladolid	735	413827N	044516W
2444	Ávila	Ávila	1130	403933N	044048W
2462	Puerto de Navacerrada	Madrid	1894	404735N	040038W
2465	Segovia	Segovia	1005	405643N	040735W
2539	Valladolid Aeropuerto	Valladolid	846	414243N	045120W
2614	Zamora	Zamora	656	413056N	054407W
2661	León Aeropuerto	León	912	423518N	053904W
2867	Salamanca Aeropuerto	Salamanca	790	405734N	052954W
2870	Salamanca	Salamanca	775	405727N	053944W
3013	Molina de Aragón	Guadalajara	1062	405030N	015244W
3129	Madrid Aeropuerto	Madrid	609	402800N	033320W
3168C	Guadalajara, El Serranillo	Guadalajara	639	403933N	031024W

3175	Torrejón de Ardoz	Madrid	607	402900N	032701W
3191E	Colmenar Viejo	Madrid	1004	404146N	034554W
3195	Madrid	Madrid	667	402443N	034041W
3196	Madrid, Cuatro Vientos	Madrid	690	402232N	034710W
3200	Getafe	Madrid	620	401800N	034321W
3260B	Toledo	Toledo	515	395305N	040243W
3469A	Cáceres	Cáceres	394	392817N	062020W
4121	Ciudad Real	Ciudad Real	628	385921N	035513W
4452	Badajoz Aeropuerto	Badajoz	185	385300N	064850W
4642E	Huelva, Ronda Este	Huelva	19	371642N	065442W
5270B	Jaén	Jaén	580	374639N	034832W
5402	Córdoba Aeropuerto	Córdoba	90	375056N	045048W
5514	Granada Base Aérea	Granada	690	370810N	033800W
5530E	Granada Aeropuerto	Granada	567	371123N	034722W
5783	Sevilla Aeropuerto	Sevilla	34	372500N	055245W
5796	Morón de la Frontera	Sevilla	87	370952N	053641W
5910	Rota, Base Aérea	Cádiz	21	363815N	061959W
5960	Jerez de la Frontera Aeropuerto	Cádiz	27	364502N	060321W
5973	Cádiz	Cádiz	2	362959N	061528W
6001	Tarifa	Cádiz	32	360050N	053556W
6155A	Málaga Aeropuerto	Málaga	16	364037N	042926W
63250	Almería Aeropuerto	Almería	21	365047N	022125W
7031	San Javier Aeropuerto	Murcia	4	374720N	004812W
7178I	Murcia	Murcia	61	380007N	011015W
7228	Alcantatilla, Base Aérea	Murcia	75	375728N	011347W
8019	Alicante/Alacant Aeropuerto	Alicante	43	381658N	003415W
8025	Alicante/Alacant	Alicante	81	382221N	002939W
8096	Cuenca	Cuenca	948	400402N	020755W
8175	Albacete Base Aérea	Albacete	702	385715N	015123W
8178D	Albacete	Albacete	674	390020N	015144W
8368U	Teruel	Teruel	900	402102N	010727W
8414A	Valencia Aeropuerto	Valencia	69	392907N	002828W
8416	Valencia	Valencia	11	392850N	002159W
8500A	Castellón de la Plana, Almazora	Castellón	43	395726N	000419W
90910	Foronda-Txokiza	Álava	513	425255N	024406W
9170	Logroño Aeropuerto	La Rioja	353	422708N	021952W
9262	Pamplona	Navarra	442	424903N	013811W
9263D	Pamplona Aeropuerto	Navarra	459	424637N	013900W
9381I	Calamocha	Teruel	890	405534N	011736W
9390	Daroca	Zaragoza	779	410652N	012436W
9434	Zaragoza Aeropuerto	Zaragoza	263	413943N	010029W
9771C	Lleida	Lleida	192	413733N	003542E
9898	Huesca Aeropuerto	Huesca	541	420500N	001935W
9981A	Tortosa	Tarragona	44	404914N	002929E

Tabla 14. Estaciones meteorológicas del periodo 1950-2011

Inhomogeneidades detectadas en las series

Temperatura máxima			Temperatura mínima					
ID	Fecha	SNHT	ID	Fecha	SNHT	ID	Fecha	SNHT
0200E	01/05/1973	103.3	0016A	01/05/2004	252.5	4121	01/01/1996	95
1484C	01/05/1976	189.2	1014	01/02/1991	95.3	4642E	01/12/1958	72.8
1690 ^a	01/12/1968	200.6	1082	01/10/2002	106.5	4642E	01/06/1984	103.4
2422	01/04/1985	93.9	1208H	01/06/1976	88	5514	01/10/1995	99
2444	01/09/1983	60.1	1208H	01/04/2001	62.8	5783	01/02/1993	381.2
2539	01/07/1999	110	1387	01/08/1984	192.5	5796	01/02/1966	100.6
2661	01/09/1987	98.6	1428	01/10/1975	97.1	5796	01/11/1995	135.1
2870	01/11/1983	102.4	1484C	01/06/1975	248.1	6001	01/12/1999	184.5
2870	01/12/2000	62.4	1495	01/09/1977	97.9	6155A	01/10/1971	109.8
3129	01/05/1999	63.4	1549	01/12/1965	128.1	6325O	01/02/1976	142.3
3129	01/12/2004	81.5	1690A	01/03/1969	83.8	7178I	01/12/1990	72.3
3168C	01/03/1971	60.9	1690A	01/09/1975	69.9	76	01/02/1973	127.7
3168C	01/07/2005	61.8	1690A	01/04/1982	117	76	01/03/1994	60.3
3175	01/12/1969	136.2	2422	01/01/1986	183.6	8019	01/05/2003	87.1
3260B	01/03/1980	293.5	2444	01/02/1983	84.7	8096	01/05/1971	97.5
367	01/12/1973	113.2	2444	01/05/1997	66.6	8096	01/06/2001	197.4
5270B	01/01/1986	335.3	2867	01/10/1998	99.3	8414A	01/02/1975	165.9
5960	01/10/1957	120.2	2870	01/12/2000	60.5	8414A	01/06/2005	85.1
6001	01/04/1972	65.7	3168C	01/04/1970	66.8	8416	01/11/1959	61.7
6325 ^e	01/06/1973	187.1	3168C	01/09/1985	64.1	8416	01/02/1966	105.8
8019	01/09/1981	91.1	3195	01/03/1985	140.4	8500A	01/01/1976	77.2
8019	01/03/1994	98.2	3196	01/03/1980	331.5	8500A	01/09/1998	168.8
8025	01/10/1971	353.7	3200	01/03/2000	245	9170	01/09/1988	214.2
9262	01/08/1963	62	3260B	01/01/1982	278.9	9262	01/03/1968	78.9
9981 ^a	01/11/1959	153.4	3469A	01/12/1982	179.2	9263D	01/10/1990	101.4
			367	01/07/1959	144	9381I	01/05/1982	96.8
			4121	01/09/1955	63.1	9434	01/10/1974	200
			4121	01/03/1963	64.4	9771C	01/03/1983	70.8
			4121	01/09/1979	76.7			

Tabla 15. Inhomogeneidades detectadas en las series del periodo 1950-2011 para la temperatura máxima y mínima

Temperatura media						Precipitaciones		
ID	Fecha	SNHT	ID	Fecha	SNHT	ID	Fecha	SNHT
0016 ^a	01/01/2008	168.3	4642E	01/06/1984	172.3	1505	01/04/1987	40
1208H	01/04/1976	65.8	5270B	01/04/1989	142.4	1549	01/11/2000	29.6
1208H	01/05/1997	141.3	5783	01/01/1990	308.6	1549	01/04/2001	29.5
1387	01/09/1987	105.4	5796	01/02/1965	117	2444	01/04/1977	45.2
1484C	01/06/1975	259.7	5796	01/07/1988	179.5	3129	01/03/1964	67.4
2444	01/02/1971	129.3	6001	01/04/1984	142.3	3168C	01/03/1965	68

2444	01/05/1999	103.1	7178I	01/10/1990	66	3469A	01/10/1976	26.2
2870	01/12/2000	73.4	8019	01/03/2003	151.9	5402	01/06/1964	33.4
3129	01/01/2000	97	8025	01/10/1970	254.1	6001	01/12/1968	25.6
3129	01/02/2005	68.2	8096	01/09/2003	120.2	6001	01/03/1972	29.7
3168C	01/10/1985	258.7	8414A	01/10/2008	148.9	6325O	01/12/2009	27.3
3195	01/03/1967	116.1	8416	01/09/1965	163.6			
3196	01/02/1970	137.4	8500A	01/01/1976	71.1			
3200	01/01/1962	137.6	8500A	01/06/1997	110.2			
367	01/02/1973	227.3	9262	01/12/1962	135.7			
4121	01/10/1979	60.9						

Tabla 16. Inhomogeneidades detectadas en las series del periodo 1950-2011 para la temperatura media y precipitaciones

Temperatura máxima			Temperatura media		
ID	Fecha	SNHT	ID	Fecha	SNHT
0372C	01/03/2016	30.6	0372C	01/05/2016	33
1475X	01/09/2016	27.1	1387E	01/05/2016	25.2
4236Y	01/02/2016	34.2	1735X	01/10/2016	25.7
5473X	01/12/2014	30.6	1738U	01/11/2016	26.3
5911A	01/03/2017	25.1	3519X	01/02/2016	28.9
8058X	01/11/2015	27.5	4236Y	01/02/2016	31.6
8072Y	01/11/2015	25.9	5473X	01/12/2014	28.3
8325X	01/09/2014	25.6	5911A	01/03/2017	25.6
8337X	01/02/2016	27.9	8058X	01/08/2015	30.9
9898	01/08/2016	28.9	8325X	01/09/2014	28.8
			8337X	01/08/2015	29.6
			8523X	01/08/2016	25.1

Tabla 17. Inhomogeneidades detectadas en las series del periodo 2014-2017 para la temperatura máxima y media

Temperatura mínima			Precipitaciones		
ID	Fecha	SNHT	ID	Fecha	SNHT
1473A	01/05/2016	26.4	1442U	01/04/2015	19.7
1735X	01/10/2016	29.9	1518A	01/04/2014	19.9
3519X	01/02/2016	30.1	2568D	01/09/2015	20.2
4148	01/02/2015	27.4	2918Y	01/05/2014	17
7012C	01/06/2016	42.3	4436Y	01/05/2014	17.2
8058X	01/08/2015	32.6	5192	01/04/2014	16.8
8523X	01/05/2016	29.2			

Tabla 18. Inhomogeneidades detectadas en las series del periodo 2014-2017 para la temperatura mínima y precipitaciones

B. Códigos

Código MySQL: Creación de la base de datos

```
/* Crear y seleccionar la base de datos del periodo 2014-2017 */  
CREATE DATABASE CLIMA;  
USE CLIMA;
```

```
/* Crear tabla de las estaciones */  
CREATE TABLE ESTACIONES (  
    Id VARCHAR(6) PRIMARY KEY,  
    Nombre VARCHAR(50) NOT NULL,  
    Altitud MEDIUMINT NOT NULL,  
    Latitud DECIMAL(4,2) NOT NULL,  
    Longitud DECIMAL(4,2) NOT NULL);
```

```
/* Insertar datos */  
LOAD DATA LOCAL INFILE 'F:/TFM/2014-2017/Estaciones.csv'  
INTO TABLE ESTACIONES  
CHARACTER SET latin1  
FIELDS TERMINATED BY ','  
LINES TERMINATED BY '\r\n'  
IGNORE 1 ROWS;
```

```
/* Crear tabla de la temperatura máxima */  
CREATE TABLE TMAX (  
    Id VARCHAR(6),  
    Fecha DATE,  
    Valor DECIMAL(3,1),  
    PRIMARY KEY (Id, Fecha));
```

```
LOAD DATA LOCAL INFILE 'H:/TFM/2014-2017/Tmax/Tmax.csv'  
INTO TABLE TMAX  
CHARACTER SET latin1  
FIELDS TERMINATED BY ','  
LINES TERMINATED BY '\r\n'  
IGNORE 1 ROWS;
```

```
/* Crear tabla de la temperatura mínima */  
CREATE TABLE TMIN (  
    Id VARCHAR(6),  
    Fecha DATE,  
    Valor DECIMAL(3,1),  
    PRIMARY KEY (Id, Fecha));
```

```
LOAD DATA LOCAL INFILE 'H:/TFM/2014-2017/Tmin/Tmin.csv'
```

```

INTO TABLE TMIN
CHARACTER SET latin1
FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;

/* Crear tabla de la temperatura media */
CREATE TABLE TMED (
    Id VARCHAR(6),
    Fecha DATE,
    Valor DECIMAL(3,1),
    PRIMARY KEY (Id, Fecha));

LOAD DATA LOCAL INFILE 'H:/TFM/2014-2017/Tmed/Tmed.csv'
INTO TABLE TMED
CHARACTER SET latin1
FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;

/* Crear tabla de las precipitaciones totales */
CREATE TABLE TPREC (
    Id VARCHAR(6),
    Fecha DATE,
    Valor DECIMAL(4,1),
    PRIMARY KEY (Id, Fecha));

LOAD DATA LOCAL INFILE 'H:/TFM/2014-2017/TPrec/TPrec.csv'
INTO TABLE TPREC
CHARACTER SET latin1
FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;

/* Crear y seleccionar la base de datos del periodo 1950-2011 */
CREATE DATABASE CLIMA;
USE CLIMA;

/* Crear tabla de las estaciones */
CREATE TABLE ESTACIONES2 (
    Id VARCHAR(6) PRIMARY KEY,
    Nombre VARCHAR(50) NOT NULL,
    Altitud MEDIUMINT NOT NULL,
    Latitud DECIMAL(4,2) NOT NULL,
    Longitud DECIMAL(4,2) NOT NULL);

LOAD DATA LOCAL INFILE 'H:/TFM/1920-2011/Estaciones.csv'

```

```
INTO TABLE ESTACIONES2
CHARACTER SET latin1
FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;
```

```
/* Crear tabla de la temperatura máxima */
CREATE TABLE TMAX2 (
    Id VARCHAR(6),
    Fecha DATE,
    Valor DECIMAL(3,1),
    PRIMARY KEY (Id, Fecha));
```

```
LOAD DATA LOCAL INFILE 'H:/TFM/1920-2011/Tmax.csv'
INTO TABLE TMAX2
CHARACTER SET latin1
FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;
```

```
/* Crear tabla de la temperatura mínima */
CREATE TABLE TMIN2 (
    Id VARCHAR(6),
    Fecha DATE,
    Valor DECIMAL(3,1),
    PRIMARY KEY (Id, Fecha));
```

```
LOAD DATA LOCAL INFILE 'H:/TFM/1920-2011/Tmin.csv'
INTO TABLE TMIN2
CHARACTER SET latin1
FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;
```

```
/* Crear tabla de la temperatura media */
CREATE TABLE TMED2 (
    Id VARCHAR(6),
    Fecha DATE,
    Valor DECIMAL(3,1),
    PRIMARY KEY (Id, Fecha));
```

```
LOAD DATA LOCAL INFILE 'H:/TFM/1920-2011/Tmed.csv'
INTO TABLE TMED2
CHARACTER SET latin1
FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;
```

```

/* Crear tabla de las precipitaciones totales */
CREATE TABLE TPREC2 (
    Id VARCHAR(6),
    Fecha DATE,
    Valor DECIMAL(4,1),
    PRIMARY KEY (Id, Fecha));

LOAD DATA LOCAL INFILE 'H:/TFM/1920-2011/TPrec.csv'
INTO TABLE TPREC2
CHARACTER SET latin1
FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;

```

Código R: Homogeneización de las series

```

library(RODBC)
library(climatol)

#TMAX

#Conexión con la base de datos de MySQL
ch <- odbcConnect("CLIMA",uid="root",pwd="root")

db2dat('Tmax',2014,2017,minny=1,daily=TRUE,ch,dformat='%Y-%m-%d',
      'TMAX','Id','Fecha','Valor','ESTACIONES',
      'Id','Nombre','Longitud','Latitud','Altitud')
#minny=1 rechaza series con menos de 1 año de datos

#Análisis exploratorio de los datos diarios originales para el control de calidad
homogen('Tmax',2014,2017,nclust=693,expl=TRUE)
#nclust=693 permite utilizar todas las series para realizar el análisis, en lugar de escoger
una muestra aleatoria de 100 series

#Agregación de los datos diarios en series mensuales
dd2m('Tmax',2014,2017)

#Homogeneización de las series mensuales
homogen('Tmax-m',2014,2017,nclust=693)

#Obtención de las series diarias homogeneizadas
homogen('Tmax',2014,2017,dz.max=13,metad=TRUE)

#Cargar archivo .rda con las series diarias homogeneizadas
load('Tmax_2014-2017.rda')

```

```

#Volcar datos a csv
dahstat('Tmax',2014,2017,stat='series')

#TMIN

ch <- odbcConnect("CLIMA",uid="root",pwd="root")
db2dat('Tmin',2014,2017,minny=1,daily=TRUE,ch,dformat='%Y-%m-%d',
      'TMIN','Id','Fecha','Valor','ESTACIONES',
      'Id','Nombre','Longitud','Latitud','Altitud')

homogen('Tmin',2014,2017,nclust=693,expl=TRUE)
dd2m('Tmin',2014,2017)
homogen('Tmin-m',2014,2017,nclust=693)
homogen('Tmin',2014,2017,dz.max=13,metad=TRUE)
load('Tmin_2014-2017.rda')
dahstat('Tmin',2014,2017,stat='series')

#TMED

ch <- odbcConnect("CLIMA",uid="root",pwd="root")
db2dat('Tmed',2014,2017,minny=1,daily=TRUE,ch,dformat='%Y-%m-%d',
      'TMED','Id','Fecha','Valor','ESTACIONES',
      'Id','Nombre','Longitud','Latitud','Altitud')

homogen('Tmed',2014,2017,nclust=693,expl=TRUE)
dd2m('Tmed',2014,2017)
homogen('Tmed-m',2014,2017,nclust=693)
homogen('Tmed',2014,2017,dz.max=10,metad=TRUE)
load('Tmed_2014-2017.rda')
dahstat('Tmed',2014,2017,stat='series')

#TPREC

ch <- odbcConnect("CLIMA",uid="root",pwd="root")
db2dat('TPrec',2014,2017,minny=1,daily=TRUE,ch,dformat='%Y-%m-%d',
      'TPREC','Id','Fecha','Valor','ESTACIONES',
      'Id','Nombre','Longitud','Latitud','Altitud')

homogen('TPrec',2014,2017,nclust=697,std=2,expl=TRUE)
dd2m('TPrec',2014,2017,valm=1)
homogen('TPrec-m',2014,2017,nclust=697,std=2)
homogen('TPrec-m',2014,2017,nclust=697,std=2,snht2=16)
homogen('TPrec',2014,2017,dz.max=32,std=2,snht2=16,metad=TRUE)
load('TPrec_2014-2017.rda')
dahstat('TPrec',2014,2017,stat='series')

```

```
#TMAX2
```

```
ch <- odbcConnect("CLIMA",uid="root",pwd="root") #conexión con la base de datos
db2dat('Tmax',1950,2011,minny=15,daily=TRUE,ch,dformat='%Y-%m-%d',
      'TMAX2','Id','Fecha','Valor','ESTACIONES2',
      'Id','Nombre','Longitud','Latitud','Altitud')
```

```
homogen('Tmax',1950,2011,expl=TRUE)
dd2m('Tmax',1950,2011)
homogen('Tmax-m',1950,2011)
homogen('Tmax-m',1950,2011,snht1=50,snht2=60)
homogen('Tmax-m',1950,2011,snht1=50,snht2=80)
homogen('Tmax-m',1950,2011,snht1=60,snht2=90)
homogen('Tmax',1950,2011,dz.max=20,snht1=60,snht2=90,metad=TRUE)
load('Tmax_1950-2011.rda')
dahstat('Tmax',1950,2011,stat='series')
```

```
#TMIN2
```

```
ch <- odbcConnect("CLIMA",uid="root",pwd="root")
db2dat('Tmin',1950,2011,minny=15,daily=TRUE,ch,dformat='%Y-%m-%d',
      'TMIN2','Id','Fecha','Valor','ESTACIONES2',
      'Id','Nombre','Longitud','Latitud','Altitud')
```

```
homogen('Tmin',1950,2011,expl=TRUE)
dd2m('Tmin',1950,2011)
homogen('Tmin-m',1950,2011)
homogen('Tmin-m',1950,2011,snht1=60,snht2=90)
homogen('Tmin',1950,2011,dz.max=12,snht1=60,snht2=90,metad=TRUE)
load('Tmin_1920-2011.rda')
dahstat('Tmin',1950,2011,stat='series')
```

```
#TMED2
```

```
ch <- odbcConnect("CLIMA",uid="root",pwd="root")
db2dat('Tmed',1950,2011,minny=15,daily=TRUE,ch,dformat='%Y-%m-%d',
      'TMED2','Id','Fecha','Valor','ESTACIONES2',
      'Id','Nombre','Longitud','Latitud','Altitud')
homogen('Tmed',1950,2011,expl=TRUE)
dd2m('Tmed',1950,2011)
homogen('Tmed-m',1950,2011)
homogen('Tmed-m',1950,2011,snht1=40,snht2=65)
homogen('Tmed-m',1950,2011,snht1=60,snht2=90)
homogen('Tmed',1950,2011,dz.max=12,snht1=60,snht2=90,metad=TRUE)
load('Tmed_1920-2011.rda')
dahstat('Tmed',1950,2011,stat='series')
```

#TPREC2

```
ch <- odbcConnect("CLIMA",uid="root",pwd="root")
db2dat('TPrec',1950,2011,minny=15,daily=TRUE,ch,dformat='%Y-%m-%d',
      'TPREC2','Id','Fecha','Valor','ESTACIONES2',
      'Id','Nombre','Longitud','Latitud','Altitud')
```

```
homogen('TPrec',1950,2011,std=2,expl=TRUE)
dd2m('TPrec',1950,2011,valm=1)
homogen('TPrec-m',1950,2011,std=2)
homogen('TPrec',1950,2011,dz.max=40,std=2,metad=TRUE)
load('TPrec_1920-2011.rda')
dahstat('TPrec',1950,2011,stat='series')
```

Código SAS Base:

Análisis de series temporales

```
/******
/******ANÁLISIS DE SERIES TEMPORALES*****
/******
```

```
LIBNAME Datos 'C:\Users\belen_000\Documents\TFM actualizado\1950-2011';
```

```
/* Agregar datos diarios a mensuales */
```

```
DATA Periodo_1950_2011;
  SET Datos.Periodo_1950_2011;
RUN;
```

```
PROC TIMESERIES DATA = Periodo_1950_2011 OUT = Datos_mensuales_media;
  ID Date INTERVAL=month ACCUMULATE=mean;
  VAR Tmax Tmin Tmed;
  BY ID;
RUN;
```

```
PROC TIMESERIES DATA = Periodo_1950_2011 OUT = Datos_mensuales_total;
  ID Date INTERVAL=month ACCUMULATE=total;
  VAR TPrec;
  BY ID;
RUN;
```

```
DATA Datos.Datos_mensuales;
  MERGE Datos_mensuales_media Datos_mensuales_total;
  FORMAT Tmax Tmin Tmed 4.1;
  FORMAT TPrec 5.1;
RUN;
```

```
%macro Estaciones(ID);
```

```
/****** ESTACION CON ID = &ID. *****
```

```
/* Transformar la variable de fecha al formato adecuado */
DATA Datos_mensuales;
```

```

SET Datos.Datos_mensuales;
Date = intnx('month', '01JAN1950'd, _n_-1);
FORMAT Date DATE.;
WHERE ID = "&ID.";
RUN;

/* Predicciones */
PROC ARIMA DATA = Datos_mensuales;
IDENTIFY VAR = Tmax(12);
ESTIMATE P=(1) Q=(12) noconstant;
FORECAST LEAD = 72 ID = Date INTERVAL = month OUT = Tmax_Predic PRINTALL;
RUN;
PROC ARIMA DATA = Datos_mensuales;
IDENTIFY VAR = Tmin(12);
ESTIMATE P=(1) Q=(12) noconstant;
FORECAST LEAD = 72 ID = Date INTERVAL = month OUT = Tmin_Predic PRINTALL;
RUN;
PROC ARIMA DATA = Datos_mensuales;
IDENTIFY VAR = Tmed(12);
ESTIMATE P=(1) Q=(12) noconstant;
FORECAST LEAD = 72 ID = Date INTERVAL = month OUT = Tmed_Predic PRINTALL;
RUN;
PROC ARIMA DATA = Datos_mensuales;
IDENTIFY VAR = TPrec(12);
ESTIMATE P=(1) Q=(12) noconstant;
FORECAST LEAD = 72 ID = Date INTERVAL = month OUT = TPrec_Predic PRINTALL;
RUN;

/*Datos de las predicciones desde 2014 hasta 2017 para comparar con los datos reales*/
DATA Tmax_Predic2;
SET Tmax_Predic;
KEEP Date Tmax;
Tmax = Forecast;
FORMAT Tmax 4.1;
WHERE Date >= '01JAN14'd;

RUN;
DATA Tmin_Predic2;
SET Tmin_Predic;
KEEP Date Tmin;
Tmin = Forecast;
FORMAT Tmin 4.1;
WHERE Date >= '01JAN14'd;

RUN;
DATA Tmed_Predic2;
SET Tmed_Predic;
KEEP Date Tmed;
Tmed = Forecast;
FORMAT Tmed 4.1;
WHERE Date >= '01JAN14'd;

RUN;
DATA TPrec_Predic2;
SET TPrec_Predic;
KEEP Date TPrec;
TPrec = Forecast;
FORMAT TPrec 5.1;
WHERE Date >= '01JAN14'd;

RUN;
DATA Predic_&ID.;
MERGE Tmax_Predic2 Tmin_Predic2 Tmed_Predic2 TPrec_Predic2;
ID = "&ID.";
RUN;

```

%mend Estaciones;

% estaciones(0016A);% estaciones(0076);% estaciones(0367);% estaciones(1082);% estaciones(1109);% estaciones(1111);% estaciones(1212E);% estaciones(1249I);% estaciones(1387);% estaciones(1387E);% estaciones(1428);% estaciones(1484C);% estaciones(1495);% estaciones(1505);% estaciones(1549);% estaciones(1690A);% estaciones(2030);% estaciones(2331);% estaciones(2422);% estaciones(2444);% estaciones(2462);% estaciones(2465);% estaciones(2614);% estaciones(2867);% estaciones(2870);% estaciones(3013);% estaciones(3129);% estaciones(3191E);% estaciones(3195);% estaciones(3260B);% estaciones(3469A);% estaciones(4121);% estaciones(4452);% estaciones(4642E);% estaciones(5270B);% estaciones(5402);% estaciones(5530E);% estaciones(5783);% estaciones(5796);% estaciones(5960);% estaciones(5973);% estaciones(6001);% estaciones(6325O);% estaciones(7178I);% estaciones(8019);% estaciones(8025);% estaciones(8096);% estaciones(8178D);% estaciones(8368U);% estaciones(8416);% estaciones(8500A);% estaciones(9170);% estaciones(9263D);% estaciones(9381I);% estaciones(9390);% estaciones(9898);

/* Unir predicciones */

DATA Datos.Predicciones;

SET Predic_0016A Predic_0076 Predic_0367 Predic_1082 Predic_1109 Predic_1111
Predic_1212E Predic_1249I Predic_1387 Predic_1387E Predic_1428 Predic_1484C
Predic_1495 Predic_1505 Predic_1549 Predic_1690A Predic_2030
Predic_2331 Predic_2422 Predic_2444 Predic_2462 Predic_2465 Predic_2614 Predic_2867
Predic_2870 Predic_3013 Predic_3129 Predic_3191E Predic_3195
Predic_3260B Predic_3469A Predic_4121 Predic_4452 Predic_4642E Predic_5270B
Predic_5402
Predic_5530E Predic_5783 Predic_5796 Predic_5960 Predic_5973
Predic_6001 Predic_6325O Predic_7178I Predic_8019 Predic_8025 Predic_8096
Predic_8178D
Predic_8368U Predic_8416 Predic_8500A Predic_9170 Predic_9263D
Predic_9381I Predic_9390 Predic_9898;

RUN;

/* COMPARACIONES DE LAS PREDICCIONES REALIZADAS CON LOS DATOS DEL PERIODO 2014-2017 */

LIBNAME DATOS2 'C:\Users\belen_000\Documents\TFM actualizado\2014-2017';

DATA Periodo_2014_2017;

SET DATOS2.Periodo_2014_2017;

RUN;

/* Agregar datos diarios a mensuales */

PROC TIMESERIES **DATA** = Periodo_2014_2017 **OUT** = Datos_mensuales_media;

ID Date **INTERVAL**=month **ACCUMULATE**=mean;

VAR Tmax Tmin Tmed;

BY ID;

RUN;

PROC TIMESERIES **DATA** = Periodo_2014_2017 **OUT** = Datos_mensuales_total;

ID Date **INTERVAL**=month **ACCUMULATE**=total;

VAR TPrec;

BY ID;

RUN;

DATA Datos_mensuales;

MERGE Datos_mensuales_media Datos_mensuales_total;

RUN;

/* Transformar la variable de fecha al formato adecuado */

DATA Datos_mensuales;

```

    SET Datos_mensuales;
    Date = intnx('month', '01JAN2014'd, _n_-1);
    FORMAT Date DATE.;
    FORMAT Tmax Tmin Tmed 4.1;
    FORMAT TPrec 5.1;
RUN;

/* Nos quedamos con los datos de las predicciones desde 2014 hasta 2017 */

DATA Predicciones;
    SET Datos.Predicciones;
    KEEP Date ID Tmax_Predic Tmin_Predic Tmed_Predic TPrec_Predic;
    Tmax_Predic = Tmax;
    Tmin_Predic = Tmin;
    Tmed_Predic = Tmed;
    TPrec_Predic = TPrec;
    FORMAT Tmax_Predic Tmin_Predic Tmed_Predic 4.1;
    FORMAT TPrec_Predic 5.1;
RUN;

DATA Datos.Comparacion;
    MERGE Datos_mensuales Predicciones;
RUN;

/* Matriz de correlaciones entre cada par de variables */

PROC CORR DATA=Datos.Comparacion OUTP=Corr_Tmax;
    VAR Tmax Tmax_Predic;
RUN;
PROC CORR DATA=Datos.Comparacion OUTP=Corr_Tmin;
    VAR Tmin Tmin_Predic;
RUN;
PROC CORR DATA=Datos.Comparacion OUTP=Corr_Tmed;
    VAR Tmed Tmed_Predic;
RUN;
PROC CORR DATA=Datos.Comparacion OUTP=Corr_TPrec;
    VAR TPrec TPrec_Predic;
RUN;

/* Gráfico para ver la relación entre cada par de variables */

PROC GPLOT DATA=Datos.Comparacion;
    PLOT Tmax*Tmax_Predic;
RUN;
PROC GPLOT DATA=Datos.Comparacion;
    PLOT Tmin*Tmin_Predic;
RUN;
PROC GPLOT DATA=Datos.Comparacion;
    PLOT Tmed*Tmed_Predic;
RUN;
PROC GPLOT DATA=Datos.Comparacion;
    PLOT TPrec*TPrec_Predic;
RUN;

```

Análisis de tendencias

```
/******  
/****** ANÁLISIS DE TENDENCIAS *****  
/******
```

```
LIBNAME Datos 'C:\Users\belen_000\Documents\TFM actualizado\1950-2011';
```

```
/* PASAR LATITUD Y LONGITUD A X-Y UTM */
```

```
%let a=6378137; *equatorial radius (default is for WGS84 datum);  
%let b=6356752.314; *polar radius (default is for WGS84 datum);  
%let k0=0.9996; *scale along central meridian;  
%let e=%sysevalf((1-(&b**2/&a**2))**0.5); *eccentricity;  
%let ep2=%sysevalf(&e**2/(1-&e**2)); *e prime squared;  
%let n=%sysevalf((&a-&b)/(&a+&b));  
%let A0=%sysevalf(&a*((1-&n)+((1-&n)*(5*&n**2)/4)+((1-&n)*(81/64*&n**4))));  
%let B0=%sysevalf((3/2*&a*&n)*((1-&n)-((1-&n)*7/8*&n**2)+(55/64*&n**4)));  
%let C0=%sysevalf((15/16*&a*&n**2)*((1-&n)+((1-&n)*(3/4*&n**2))));  
%let D0=%sysevalf((35/48*&a*&n**3)*((1-&n)+(11/16*&n**2)));  
%let E0=%sysevalf((315/51*&a*&n**4)*(1-&n));
```

```
data datos.variables_estaciones (drop=LM deltalon latr lonr rho nu MA Ki Kii Kiii Kiv Kv A6 RN  
hemisphere zone);
```

```
set datos.variables_estaciones;
```

```
if LATITUD < 0 then hemisphere = "S"; /**variable que contiene la latitud*/
```

```
else hemisphere = "N";
```

```
zone=31+floor(LONGITUD/6);
```

```
LM=6*zone-183;
```

```
deltalon=(longitud-LM)*constant('pi')/180;
```

```
latr = LATITUD*constant('pi')/180;
```

```
lonr = LONGITUD*constant('pi')/180;
```

```
rho = &a*(1-&e**2)/((1-(&e*sin(latr))**2)**1.5);
```

```
nu = &a/((1-(&e*sin(latr))**2)**0.5);
```

```
MA = (&A0*latr)-(&B0*sin(2*latr))+(&C0*sin(4*latr))-(&D0*sin(6*latr))+(&E0*sin(8*latr));
```

```
Ki = MA*&k0;
```

```
Kii = nu*sin(latr)*cos(latr)/2;
```

```
Kiii = ((nu*sin(latr)*cos(latr)**3)/24)*(5-
```

```
tan(latr)**2+9*&ep2*cos(latr)**2+4*&ep2**2*cos(latr)**4)*&k0;
```

```
Kiv = nu*cos(latr)*&k0;
```

```
Kv = cos(latr)**3*nu/6*(1-tan(latr)**2+&ep2*cos(latr)**2)*&k0;
```

```
A6 = ((deltalon)**6*nu*sin(latr)*cos(latr)**5/720)*(61-
```

```
58*tan(latr)**2+tan(latr)**4+270*&ep2*cos(latr)**2-330*&ep2*sin(latr)**2)*&k0;
```

```
RN =(Ki+Kii*deltalon**2+Kiii*deltalon**4);
```

```
if RN < 0 then y=10000000+RN;
```

```
else y=RN;
```

```
x = 500000+(Kiv*deltalon+Kv*deltalon**3);
```

```
run;
```

```
/* Crear variables Mes y Year2 */
```

```
DATA Datos_mensuales;
```

```
SET Datos.Datos_mensuales;
```

```
Mes2 = substr(PUT(Date, MONYY7.), 1, 3);
```

```
Year3 = substr(PUT(Date, MONYY7.), 4, 4);
```

```
RUN;
```

```
DATA Datos_mensuales;
```

```
SET Datos_mensuales;
```

```

IF Mes2 = 'JAN' THEN Mes = 1;
ELSE IF Mes2 = 'FEB' THEN Mes = 2;
ELSE IF Mes2 = 'MAR' THEN Mes = 3;
ELSE IF Mes2 = 'APR' THEN Mes = 4;
ELSE IF Mes2 = 'MAY' THEN Mes = 5;
ELSE IF Mes2 = 'JUN' THEN Mes = 6;
ELSE IF Mes2 = 'JUL' THEN Mes = 7;
ELSE IF Mes2 = 'AUG' THEN Mes = 8;
ELSE IF Mes2 = 'SEP' THEN Mes = 9;
ELSE IF Mes2 = 'OCT' THEN Mes = 10;
ELSE IF Mes2 = 'NOV' THEN Mes = 11;
ELSE IF Mes2 = 'DEC' THEN Mes = 12;
Year = Year3 * 1;
Year2 = Year3 - 1950;

RUN;

PROC SQL;
CREATE TABLE Datos.Datos_mensuales2 AS SELECT DISTINCT
    ID, Date, Year, Year2, Mes, Tmax, Tmin, Tmed, TPrec
FROM Datos_mensuales;
QUIT;

/* PREBLANQUEO Y EXAMEN DE LA TENDENCIA TMAX */

/* Cálculo de la correlación de orden 1 en caso de que exista: Test de Durbin-Watson */

PROC SORT DATA = Datos.Datos_mensuales2 OUT = Datos_mensuales;
BY ID Mes;

RUN;

ods output DWTest = Durbin ARParameterEstimates = AR;
PROC AUTOREG DATA = Datos_mensuales OUTEST = pp dw = 1 dwprob; /*dw = p: nº
periodos en un año: 1*/
BY ID Mes;
MODEL Tmax = / NLAG = 1;

RUN;

DATA Durbin;
SET Durbin;
RETAIN p_autocor;
BY ID Mes;
IF FIRST.Mes THEN p_autocor = probDW;
IF LAST.Mes THEN DO;
    p_autocor1 = probdw;
OUTPUT;
END;
DROP ProbDWNeg ProbDW model order dw;

RUN;

DATA Union;
MERGE Durbin AR;
BY ID Mes;
DROP model;
RENAME estimate = r1;

RUN;

/* Preblanqueo para aquellos casos en los que sea necesario */

DATA MEDIAMES;
MERGE Datos_mensuales Union;

```

```

BY ID Mes;
IF p_autocor < 0.05 THEN Tmax_ = (Tmax - r1 * LAG(Tmax)) / (1 - r1);
ELSE Tmax_ = Tmax;
IF FIRST.Mes = 1 THEN Tmax_ = Tmax;
RUN;

```

```

/* CONTRASTES DE MAN-KENDALL Y THEIL-SEN */

```

```

/* Macro de Mann-Kendall */

```

```

%macro mktrend(data= ,var= ,out= );
proc means data=&data noprint;
var &var;
output out=_temp n=n;
run;
data _null_;
set _temp;
call symput("n",trim(left(n)));
run;
proc transpose data=&data out=&out;
run;
data &out;
keep s vars;
set &out;
array c{&n} col1-col&n;
array u{&n} uu1-uu&n;
temp=upcase("&var");
if upcase(_name_) = temp then do;
s=0;
do i=1 to &n;
if i=1 then u[1]=0;
do j=i to &n;
if i=1 then u[j]=0;
if i<j then do;
signum=sign(c[j]-c[i]);
if signum=0 then do; *tie*;
if u[i]>=0 then do;
u[i]=u[i]+1;
u[j]=-i;
end;
end;
s=s+signum;
end;
end;
j=0;
vars=0;
do i=1 to &n;
u[i]=(1+u[i])*(u[i]>0); * multiplicity of a tie *;
j=j+(u[i]>0); *j=number of ties *;
if u[i]>0 then do;
put c[i] ' has multiplicity ' u[i] '.';
u[j]=u[i];
vars=vars+(u[i])*(u[i]-1)*(2*u[i]+5);
end;
end;
vars=-vars+(&n)*(&n-1)*(2*&n+5);
vars=vars/18;
if j<&n then u[j+1]=0;
output;

```

```

*do i=0 to j;
*if i=0 then put 'There are ' j 'sets of ties.';
*end;
end;
run;
data &out;
set &out;
check=s/sqrt(vars);
p=1-probnorm(check);
RUN;
/**
proc print data=&out;
title 'Mann Kendall Trend Test';
title2 "&n Observations";
run;
**/
%mend;

%macro RKENDALL (DATA=, VAR=, OUT=, BY=, seasonal=, VARSEASONAL=, BYTEMP=,
YEAR=, alfanumericos=);
proc sort data=&data; by &by &year &varseasonal;
options nonotes;
quit;
ODS OUTPUT OneWayFreqs=FRIQUI1;
PROC FREQ DATA=&DATA; TABLES &BY / NOPERCENT NOCUM;RUN;
ODS OUTPUT CLOSE;
PROC IML;
ALA1={&VAR};
ALA=T(ALA1);
NVAR=NROW(ALA);
CREATE ALA VAR {ALA}; /*ALA contiene a las variables*/
APPEND;
CLOSE ALA;
CREATE NVAR VAR {NVAR}; /*NVAR contiene el número de variables*/
APPEND;
CLOSE NVAR;
USE FRIQUI1;
READ ALL VAR {&BY} INTO PUNTITOS; /*PUNTITOS es una matriz con las
estaciones*/
NPUNTOS=NROW(PUNTITOS);
CREATE PUNTILLOS VAR {NPUNTOS}; /*PUNTILLOS C.datos que contiene
número de estaciones*/
APPEND;
CLOSE PUNTILLOS;
QUIT;
DATA _NULL_; SET PUNTILLOS; call symput("MI_N",trim(left(nPUNTOS)));
DATA _NULL_; SET NVAR; call symput("MI_X",trim(left(nVAR)));
%if &seasonal=1 %then %do;
ODS OUTPUT OneWayFreqs=FRIQUI2;
PROC FREQ DATA=&DATA; TABLES &varseasonal/NOPERCENT
NOCUM;RUN; /*FRIQUI2 contiene el nombre de los periodos*/
ODS OUTPUT CLOSE;
proc iml; use friqui2; read all into oo; ns=nrow(oo);
create ns var{ns}; append ; close ns;
quit;
data _null_; set ns;
call symput("nseason",trim(left(ns))); /*NSEASON es el número de
periodos*/
run;
%end;
%DO MI=1 %TO &MI_N;

```

```

DATA _NULL_; REFER=&MI; SET FRIQUI POINT=REFER;
  call symput("REGISTRO",trim(left(&BY)));
  STOP;
RUN;
%DO mu=1 %to &mi_x;
  DATA _NULL_; REFER=&mu; set ala point=refer;
  call symput("Rvar",trim(left(ala))); /*Rvar es la variable que se
examina*/
  stop;
run;
%if &seasonal=1 %then %do;
  %do MI_EN=1 %TO &NSEASON;
    data _null_ ; refer=&MI_EN;
    SET FRIQUI2 POINT=REFER;
    CALL SYMPUT
("PRIMAVERA",trim(left(&VARSEASONAL)));
  STOP;
  RUN;
  %if &alfanumericos=11 %then %do;
    data basura; set &data;
    where &by="&registro" AND
&VARSEASONAL="&primavera"; /*by y varseasonal son alfanuméricas*/
    keep &by &rvar &year &bytemp
&VARSEASONAL;
    run;
  %end;
  %if &alfanumericos=10 %then %do;
    data basura; set &data;
    where &by="&registro" AND
&VARSEASONAL=&primavera; /*by es alfanumérica*/
    keep &by &rvar &year &bytemp
&VARSEASONAL;
    run;
  %end;
  %if &alfanumericos=01 %then %do;
    data basura; set &data;
    where &by=&registro AND
&VARSEASONAL="&primavera"; /*varseasonal es alfanumérica*/
    keep &by &rvar &year &bytemp
&VARSEASONAL;
    run;
  %end;
  %if &alfanumericos=0 %then %do;
    data basura; set &data;
    where &by=&registro AND
&VARSEASONAL=&primavera; /*by y varseasonal son numéricas*/
    keep &by &rvar &year &bytemp
&VARSEASONAL;
    run;
  %end;
  %mktrend(data=BASURA, var=&RVAR ,out=OUT_EN );
  DATA OUT_EN; SET OUT_EN; SEASONAL=&MI_EN;
  RUN;
  %IF &MI_EN=1 %THEN %DO;
    DATA OUT1; SET OUT_EN; RUN;
  %END;
  %ELSE %DO;
    DATA OUT1; SET OUT1 OUT_EN; RUN;
  %END;
%END;
%END;

```

```

        %ELSE %DO;
            %if &alfanumericos=1 %then %do;
                data basura; set &data;
                where &by="&registro"; /*by es alfanumérica*/
                keep &by &Rvar &year &bytemp;
                run;
            %end;
            %if &alfanumericos=0 %then %do;
                data basura; set &data;
                where &by=&registro;
                keep &by &Rvar &year &bytemp; /*by es numérica*/
                run;
            %end;
            %mktrend(data=BASURA, var=&RVAR ,out=OUT1);
        %END;
    DATA OUT1; SET OUT1; informat mivariable $20.;MIVARIABLE="&RVAR";
format mivariable $20.; RUN;
    %IF &MU=1 %THEN %DO;
        DATA OUT2; SET OUT1; format mivariable $20.;RUN;
        %END;
    %ELSE %DO;
        DATA OUT2; SET OUT2 OUT1; format mivariable $20.;
informat mivariable $20.; RUN;
    %END;
%END;
DATA OUT2; SET OUT2; MIPUNTO="&REGISTRO";format mipunto $20.; RUN;
%IF &MI=1 %THEN %DO;
    DATA &OUT; SET OUT2; RUN;
    %END;
%ELSE %DO;
    DATA &OUT; SET &OUT OUT2; RUN;
    %END;
%END;
options notes;
PROC DATASETS;
    DELETE ALA FRIQUI FRIQUI2 OUT1 OUT2 OUT_EN PUNTILLOS;
RUN;
proc print data=&out;run;
%MEND;

```

```

%RKENDALL (DATA = MEDIAMES, VAR = Tmax_, OUT = Kendall, BY = ID, seasonal = 1,
VARSEASONAL = Mes, BYTEMP =, YEAR = Year, alfanumericos = 10);
/* seasonal = 1: datos para cada mes de forma diferenciada (porque los datos tienen
estacionalidad)
alfanumericos = 10: ID es alfanumerica (1) y Mes es numerica (0) */

```

```

DATA Kendall2;
    SET Kendall;
    IF p < 0.05 THEN Tendencia = 'Si';
        ELSE Tendencia = 'No';
    DROP s vars check p;
    RENAME seasonal = Mes mipunto = ID;
RUN;

```

```

/* Contraste de Theil-Sen; macro Theil para estaciones y meses */

```

```

%macro theil (data=, time=, var=);
proc sort data=&data; by &time; run;
proc iml; use &data; read all into lista; n=nrow(lista);

```

```

        create eln from n;
        append from n ;
        close eln;
quit;
data _null_ ; set eln; call symput ("nn",col1); %put &nn; run;
proc transpose data=&data out=traspose;
    var &var; run;
    data traspose2; set traspose;
        array col (&nn); z=&nn-1;
        do i=1 to z; do j=i+1 to &nn;
            b= (col(j)-col(i))/(j-i);
                output;
                keep b;
            end; end;
proc means data=traspose2 noprint;
var b;
output out=mediana median=Beta;
run;
proc means data=&data noprint; var &var; output out=media mean=alfa;
run;
data resul; merge media mediana; keep alfa beta; run;
proc print data=resul;run;
%mend;

%macro RTHAIL(DATA=, NBY=, BY1=, BY2=, TIME=, VAR=);
    ODS OUTPUT OneWayFreqs=F1;
    PROC FREQ DATA=&DATA;
    TABLES &BY1 ;
    RUN;
    ODS OUTPUT CLOSE;
    DATA F1; SET F1; KEEP &BY1;
    PROC IML; USE F1; READ ALL VAR {&BY1} INTO FF1; MY_N1=NROW(FF1);
        CREATE MY_N1 var{my_n1};
        append;
        close my_n1;
    QUIT;
    %IF &NBY>1 %THEN %DO;
        ODS OUTPUT OneWayFreqs=F2;
        PROC FREQ DATA=&DATA;
        TABLES &BY2 ;
        RUN;
        ODS OUTPUT CLOSE;
        DATA F2; SET F2; KEEP &BY2;
    PROC IML; USE F2; READ ALL INTO FF2; MY_N2=NROW(FF2);
        CREATE MY_N2 var{my_n2};
        append;
        close my_n2;
    %END;
    DATA _NULL_; SET MY_n1; call symput("n_est",trim(left(my_n1))); %put &n_est;
run; /* N_est= número de estaciones*/
    DATA _NULL_; SET my_n2; call symput("n_mes",trim(left(my_n2))); %put &n_mes;
run; /* N_mes= número de meses*/
    %put 'paso por aqui';
    %do rs=1 %to &N_est;
        %do tw=1 %to &N_mes;
    data _null_ ; referencia1=&rs;
        set f1 point=referencia1;
        call symput ("laesta",&by1); stop; run;
    data _null_ ; referencia2=&tw; set f2 point=referencia2;
        call symput ("elmes", &by2); stop; run;
    data trabajo; set &data;

```

```

where &by1="&laesta" and &by2=&elmes; /*modificar según sea carácter o no*/
run;
%theil (data=trabajo, time=&time , var=&var);
data resul; set resul; &by1="&laesta" ; &by2=&elmes;
%IF &rs=1 and &tw=1 %then %do;
    data resultado_thail; set resul; run;
    %end;
%else %do;
    data resultado_thail; set resultado_thail resul;
run;
%end;
%end;
%end;
proc print data=resultado_thail; format alfa 6.2 beta 6.3; run;
%mend;

```

```

%RTHAIL (DATA = MEDIAMES, NBY = 2, BY1 = ID, BY2 = Mes, TIME = Year2, VAR = Tmax_);

```

```

/* Calcula 1 tendencia por cada mes y estación: 672 series */
/* NBY = 2: número de variables por las que segmentamos (ID y Mes)
   TIME = Year2: el año 1950 va a ser el año 0 para que el valor no sea muy negativo, para que
   el valor de alfa tenga sentido */

```

```

/* Si la serie tiene tendencia --> asignar valores alfa y beta
   Si la serie no tiene tendencia --> asignar valor alfa (beta = 0) */

```

```

PROC SQL;
    CREATE TABLE Tendencia AS SELECT DISTINCT
        A.ID, A.Mes, A.Tendencia, B.Alfa, B.Beta,
        CASE
            WHEN A.Tendencia = 'Si' THEN B.Beta
            WHEN A.Tendencia = 'No' THEN 0
        END AS Beta2
    FROM Kendall2 A LEFT JOIN Resultado_thail B
    ON A.ID = B.ID AND A.Mes = B.Mes;

```

```

QUIT;

```

```

/* Unir los datos calculados de la tendencia y las variables explicativas de las estaciones */

```

```

PROC SQL;
    CREATE TABLE Datos.Tendencias_Estaciones_Tmax AS SELECT DISTINCT
        A.ID, A.Mes, A.Alfa, A.Beta2 AS Beta,
        B.Altitud, B.X, B.Y, B.DIST_MAR, B.MAR
    FROM Tendencia A LEFT JOIN Datos.Variables_estaciones B
    ON A.ID = B.ID;

```

```

QUIT;

```

```

/* PREBLANQUEO Y EXAMEN DE LA TENDENCIA TMIN */

```

```

/* Cálculo de la correlación de orden 1 si es necesaria: Test de Durbin-Watson */

```

```

PROC SORT DATA = Datos.Datos_mensuales OUT = Datos_mensuales;
    BY ID Mes;

```

```

RUN;

```

```

ods output DWTest = Durbin ARParameterEstimates = AR;

```

```

PROC AUTOREG DATA = Datos_mensuales OUTEST = pp dw = 1 dwprob; /*dw = p: n°
periodos en un año: 1*/

```

```

        BY ID Mes;
        MODEL Tmin = / NLAG = 1;
RUN;

DATA Durbin;
    SET Durbin;
    RETAIN p_autocor;
    BY ID Mes;
    IF FIRST.Mes THEN p_autocor = probdW;
    IF LAST.Mes THEN DO;
        p_autocor1 = probdW;
    OUTPUT;
    END;
    DROP ProbDWNeg ProbDW model order dw;
RUN;

DATA Union;
    MERGE Durbin AR;
    BY ID Mes;
    DROP model;
    RENAME estimate = r1;
RUN;

/* Preblanqueo para aquellos casos en los que sea necesario */

DATA MEDIAMES;
    MERGE Datos_mensuales Union;
    BY ID Mes;
    IF p_autocor < 0.05 THEN Tmin_ = (Tmin - r1 * LAG(Tmin)) / (1 - r1);
    ELSE Tmin_ = Tmin;
    IF FIRST.Mes = 1 THEN Tmin_ = Tmin;
RUN;

/* Contrastes de Man-Kendall y Theil-Send */

%RKENDALL (DATA = MEDIAMES, VAR = Tmin_, OUT = Kendall, BY = ID, seasonal = 1,
VARSEASONAL = Mes, BYTEMP =, YEAR = Year, alfanumericos = 10);

DATA Kendall2;
    SET Kendall;
    IF p < 0.05 THEN Tendencia = 'Si';
    ELSE Tendencia = 'No';
    DROP s vars check p;
    RENAME seasonal = Mes mipunto = ID;
RUN;

%RTHAIL (DATA = MEDIAMES, NBY = 2, BY1 = ID, BY2 = Mes, TIME = Year2, VAR = Tmin_);

PROC SQL;
    CREATE TABLE Tendencia AS SELECT DISTINCT
        A.ID, A.Mes, A.Tendencia, B.Alfa, B.Beta,
        CASE
            WHEN A.Tendencia = 'Si' THEN B.Beta
            WHEN A.Tendencia = 'No' THEN 0
        END AS Beta2
    FROM Kendall2 A LEFT JOIN Resultado_thail B
    ON A.ID = B.ID AND A.Mes = B.Mes;
QUIT;

/* Unir los datos calculados de la tendencia y las variables explicativas de las estaciones */

```

```

PROC SQL;
    CREATE TABLE Datos.Tendencias_Estaciones_Tmin AS SELECT DISTINCT
        A.ID, A.Mes, A.Alfa, A.Beta2 AS Beta,
        B.Altitud, B.X, B.Y, B.DIST_MAR, B.MAR
    FROM Tendencia A LEFT JOIN Datos.Variables_estaciones B
    ON A.ID = B.ID;
QUIT;

%REGBUENA (DATA = Datos.Tendencias_Estaciones_Tmin, BY = ID, VDEP = Beta, VIND =
Altitud X Y DIST_MAR, resultado = REG_Tendencia);

/* PREBLANQUEO Y EXAMEN DE LA TENDENCIA */

/* Cálculo de la correlación de orden 1 si es necesaria: Test de Durbin-Watson */

PROC SORT DATA = Datos.Datos_mensuales OUT = Datos_mensuales;
    BY ID Mes;
RUN;

ods output DWTest = Durbin ARParameterEstimates = AR;
PROC AUTOREG DATA = Datos_mensuales OUTEST = pp dw = 1 dwprob; /*dw = p: nº
periodos en un año: 1*/
    BY ID Mes;
    MODEL Tmed = / NLAG = 1;
RUN;

DATA Durbin;
    SET Durbin;
    RETAIN p_autocor;
    BY ID Mes;
    IF FIRST.Mes THEN p_autocor = probdW;
    IF LAST.Mes THEN DO;
        p_autocor1 = probdW;
    OUTPUT;
    END;
    DROP ProbDWNeg ProbDW model order dw;
RUN;

DATA Union;
    MERGE Durbin AR;
    BY ID Mes;
    DROP model;
    RENAME estimate = r1;
RUN;

/* Preblanqueo para aquellos casos en los que sea necesario */

DATA MEDIAMES;
    MERGE Datos_mensuales Union;
    BY ID Mes;
    IF p_autocor < 0.05 THEN Tmed_ = (Tmed - r1 * LAG(Tmed)) / (1 - r1);
    ELSE Tmed_ = Tmed;
    IF FIRST.Mes = 1 THEN Tmed_ = Tmed;
RUN;

/* Contrastes de Man-Kendall y Theil-Send */

%RKENDALL (DATA = MEDIAMES, VAR = Tmed_, OUT = Kendall, BY = ID, seasonal = 1,
VARSEASONAL = Mes, BYTEMP =, YEAR = Year, alfanumericos = 10);

```

```

DATA Kendall2;
  SET Kendall;
  IF p < 0.05 THEN Tendencia = 'Si';
  ELSE Tendencia = 'No';
  DROP s vars check p;
  RENAME seasonal = Mes mipunto = ID;
RUN;

%RTHAIL (DATA = MEDIAMES, NBY = 2, BY1 = ID, BY2 = Mes, TIME = Year2, VAR =
Tmed_);

PROC SQL;
  CREATE TABLE Tendencia AS SELECT DISTINCT
    A.ID, A.Mes, A.Tendencia, B.Alfa, B.Beta,
    CASE
      WHEN A.Tendencia = 'Si' THEN B.Beta
      WHEN A.Tendencia = 'No' THEN 0
    END AS Beta2
  FROM Kendall2 A LEFT JOIN Resultado_thail B
  ON A.ID = B.ID AND A.Mes = B.Mes;
QUIT;

/* Unir los datos calculados de la tendencia y las variables explicativas de las estaciones */

PROC SQL;
  CREATE TABLE Datos.Tendencias_Estaciones_Tmed AS SELECT DISTINCT
    A.ID, A.Mes, A.Alfa, A.Beta2 AS Beta,
    B.Altitud, B.X, B.Y, B.DIST_MAR, B.MAR
  FROM Tendencia A LEFT JOIN Datos.Variables_estaciones B
  ON A.ID = B.ID;
QUIT;

%REGBUENA (DATA = Datos.Tendencias_Estaciones_Tmed, BY = ID, VDEP = Beta, VIND =
Altitud X Y DIST_MAR, resultado = REG_Tendencia);

/* PREBLANQUEO Y EXAMEN DE LA TENDENCIA */

/* Cálculo de la correlación de orden 1 si es necesaria: Test de Durbin-Watson */

PROC SORT DATA = Datos.Datos_mensuales OUT = Datos_mensuales;
  BY ID Mes;
RUN;

ods output DWTest = Durbin ARParameterEstimates = AR;
PROC AUTOREG DATA = Datos_mensuales OUTEST = pp dw = 1 dwprob; /*dw = p: nº
periodos en un año: 1*/
  BY ID Mes;
  MODEL Tprec = / NLAG = 1;
RUN;

DATA Durbin;
  SET Durbin;
  RETAIN p_autocor;
  BY ID Mes;
  IF FIRST.Mes THEN p_autocor = probdW;
  IF LAST.Mes THEN DO;
    p_autocor1 = probdw;
  OUTPUT;

```

```

        END;
        DROP ProbDWNeg ProbDW model order dw;
RUN;

DATA Union;
    MERGE Durbin AR;
    BY ID Mes;
    DROP model;
    RENAME estimate = r1;
RUN;

/* Preblanqueo para aquellos casos en los que sea necesario */

DATA MEDIAMES;
    MERGE Datos_mensuales Union;
    BY ID Mes;
    IF p_autocor < 0.05 THEN Tprec_ = (Tprec - r1 * LAG(Tprec)) / (1 - r1);
        ELSE Tprec_ = Tprec;
    IF FIRST.Mes = 1 THEN Tprec_ = Tprec;
RUN;

/* Contrastes de Man-Kendall y Theil-Send */

%RKENDALL (DATA = MEDIAMES, VAR = Tprec_, OUT = Kendall, BY = ID, seasonal = 1,
VARSEASONAL = Mes, BYTEMP = , YEAR = Year, alfanumericos = 10);

DATA Kendall2;
    SET Kendall;
    IF p < 0.05 THEN Tendencia = 'Si';
        ELSE Tendencia = 'No';
    DROP s vars check p;
    RENAME seasonal = Mes mipunto = ID;
RUN;

/* Unir los valores de alfa y beta de las tendencias de las temperaturas máximas, mínimas y medias */

PROC SQL;
    CREATE TABLE Datos.Tendencias_estaciones AS SELECT DISTINCT
        A.ID, A.Mes, A.Alfa AS Alfa_Tmax, A.Beta AS Beta_Tmax,
        B.Alfa AS Alfa_Tmin, B.Beta AS Beta_Tmin,
        C.Alfa AS Alfa_Tmed, C.Beta AS Beta_Tmed,
        A.Altitud, A.X, A.Y, A.Dist_Mar, A.Mar
    FROM Datos.Tendencias_estaciones_Tmax A LEFT JOIN
    Datos.Tendencias_estaciones_Tmin B
        ON A.ID = B.ID AND A.Mes = B.Mes
    LEFT JOIN Datos.Tendencias_estaciones_Tmed C
        ON A.ID = C.ID AND A.Mes = C.Mes;
QUIT;

PROC SQL;
    CREATE TABLE Datos.Tendencias_estaciones AS SELECT DISTINCT
        A.*,
        B.Provincia
    FROM Datos.Tendencias_estaciones A LEFT JOIN TFM.Nombres B
        ON A.ID = B.ID;
QUIT;

```

```
/* Macro regresión lineal */
```

```
%MACRO REGBUENA (DATA=, BY=, VDEP=, VIND=, resultado=, WHERE=);  
  proc sort data=&data; by &by; run;  
  ods output DWStatistic = midw;  
    ods noresults;  
    ods listing close;  
  Proc reg data=&DATA outest=&resultado edf;  
    by &by ;  
    &WHERE ;  
    MODEL &VDEP=&VIND /* selection=forward dwprob*/;  
quit;  
  ODS OUTPUT CLOSE;  
  ods listing;  
  ods results;  
  data midw; set midw; if label1='Pr < DW' then label1='DWpos';  
    if label1='Pr > DW' THEN LABEL1='DWneg'; run;  
  proc sort data=midw; by dependent id;run;  
  proc transpose data=midw out=midw2;  
  var nvalue1;  
  id label1;  
  by dependent &by;  
  run;  
  data midw2; set midw2; drop Number_of_Observations; rename dependent=_depvar_;  
run;  
  data &resultado; set &resultado; drop _type_ _model_; run;  
  proc sort data=&resultado; by _depvar_ &by; run;  
  proc sort data=midw2; by _depvar_ &by; run;  
  data &resultado; merge &resultado midw2; by _depvar_ &by; run;  
  proc print data=&resultado; run;  
%mend;
```

```
%REGBUENA (DATA = Datos.Tendencias_Estaciones, BY = Mes, VDEP = Beta_Tmax  
Beta_Tmin Beta_Tmed, VIND = Altitud X Y DIST_MAR, resultado = Regresion_Tendencia);
```

```
DATA Datos.Regresion_Tendencia;  
  SET Regresion_Tendencia;  
RUN;
```

```
/* MAPAS DE LAS TENDENCIA DE LAS TEMPERATURAS */
```

```
PROC MAPIMPORT DATAFILE = "C:\Users\belen_000\Documents\TFM actualizado\Archivos  
tutor\ESP_adm_shp\esp_adm2.shp"
```

```
  OUT = ESPANIA;  
RUN;
```

```
PROC SORT DATA = ESPANIA NODUPKEY OUT = ESPANIA2;  
  BY ID_2;  
  WHERE Name_1 NOT IN ('Islas Baleares', 'Islas Canarias', 'Ceuta y Melilla');  
RUN;
```

```
PROC SQL;  
  CREATE TABLE ESPANIA3 AS SELECT DISTINCT  
    A.*,  
    B.Mes, B.Beta_Tmax, B.Beta_Tmin, B.Beta_Tmed  
  FROM ESPANIA2 A LEFT JOIN Datos.Tendencias B  
  ON A.Name_2 = B.Provincia;  
RUN;
```

```

DATA ESPANIA4;
  SET ESPANIA3;
  DROP Beta_Tmax Beta_Tmin Beta_Tmed;
  POP = Beta_Tmax;
  Temperatura = 'Máxima';
RUN;

DATA ESPANIA5;
  SET ESPANIA3;
  DROP Beta_Tmax Beta_Tmin Beta_Tmed;
  POP = Beta_Tmin;
  Temperatura = 'Mínima';
RUN;

DATA ESPANIA6;
  SET ESPANIA3;
  DROP Beta_Tmax Beta_Tmin Beta_Tmed;
  POP = Beta_Tmed;
  Temperatura = 'Media';
RUN;

DATA ESPANIA7;
  SET ESPANIA4 ESPANIA5 ESPANIA6;
RUN;

PROC SORT DATA = ESPANIA7;
  BY Temperatura Mes;
RUN;

goptions reset=all border;
TITLE1 "TENDENCIA DE LA TEMPERATURA EN ESPAÑA";
TITLE2 "1950 - 2011";
LEGEND1 LABEL = ("PENDIENTE");

ods path(prepend) work.templat(update);

proc template;
  define style styles.colorramp;
    parent=styles.default;
    style twocolorramp / startcolor=WHITE endcolor=RED;
    style graphdata1 from graphdata1 / color=WHITE;
    style graphdata2 from graphdata2 / color=RED;
  end;
run;

ods listing style=styles.colorramp;
ods html file='colorramp.html' path='.' style=styles.colorramp;

PROC GMAP DATA = ESPANIA7 MAP = ESPANIA;
  ID ID_2;
  CHORO POP / UNIFORM
  LEGEND = LEGEND1;
  BY Temperatura Mes;
RUN; QUIT;

```