

CONTROL DE CALIDAD Y HOMOGENEIZACIÓN DE LOS DATOS OBSERVADOS SUMINISTRADOS

PARA EL PROYECTO

**“ELABORACIÓN DE ESCENARIOS LOCALES Y REGIONALES
DE CAMBIO CLIMÁTICO ADAPTADOS AL SEXTO INFORME
DEL GRUPO INTERGUBERNAMENTAL DE EXPERTOS SOBRE
EL CAMBIO CLIMÁTICO (IPCC) Y SU DIFUSIÓN A TRAVÉS
DE LAS TECNOLOGÍAS DE LA INFORMACIÓN Y LA
COMUNICACIÓN”**



CONTROL DE CALIDAD Y HOMOGENEIZACIÓN DE LOS DATOS OBSERVADOS SUMINISTRADOS	0
INTRODUCCIÓN	2
1. DATOS METEOROLÓGICOS SUMINISTRADOS	3
2. FUNDAMENTO TEÓRICO	8
2.1. CONTROL DE CALIDAD	8
2.1.1. TEMPERATURA	8
COHERENCIA BÁSICA	8
OUTLIERS	9
2.1.2. PRECIPITACIÓN	10
COHERENCIA BÁSICA	10
OUTLIERS	10
ELIMINACIÓN DE CEROS CONSECUTIVOS	11
2.2. HOMOGENEIZACIÓN	11
2.2.1. HOMOGENEIZACIÓN: EJEMPLO TEÓRICO	12
2.2.2. EJEMPLOS	14
3. TEMPERATURA	19
3.1. CONTROL DE CALIDAD	19
3.1.1. COHERENCIA BÁSICA	19
3.1.2. OUTLIERS. VALORES EXTREMOS.	19
3.1.3. OUTLIERS. LÍMITE DISTRIBUCIÓN MENSUAL.	22
3.2. HOMOGENEIZACIÓN	22
4. PRECIPITACIÓN	30
4.1 CONTROL DE CALIDAD	30
4.1.1. COHERENCIA BÁSICA	30
4.1.2. ELIMINACIÓN DE OUTLIERS	30
4.1.3. ELIMINACIÓN DE CEROS	32
4.2. HOMOGENEIZACIÓN	33

INTRODUCCIÓN

El presente documento detalla el proceso de Control de calidad y Homogeneización de los datos observados suministrados para el proyecto “Elaboración de escenarios locales y regionales de cambio climático adaptados al sexto informe del grupo intergubernamental de expertos sobre el cambio climático (IPCC) y su difusión a través de las tecnologías de la información y la comunicación” de la Junta de Andalucía.

Cuando se reciben series temporales de datos meteorológicos observados siempre cabe la posibilidad de que algunos de tales datos sean falsos o erróneos, debido o bien a errores del observador si es un observatorio meteorológico manual o a problemas de conexión si es un observatorio meteorológico automático - además de los propios posibles problemas de los equipos de medición por sí mismos. Es por eso por lo que, antes de empezar a utilizar tales series de datos observados, debe realizarse un control de calidad de tales datos para determinar cuánto de fiables son éstos y qué actuaciones se pueden llevar a cabo.

Lo que sigue es la sistematización de tal control de calidad para las variables meteorológicas de precipitación y temperatura diarias, referido a los valores asociados a los observatorios meteorológicos usados en este proyecto.

1. DATOS METEOROLÓGICOS

SUMINISTRADOS

Para este proyecto se han suministrado un total de 2.463 observatorios meteorológicos con datos de precipitación y 1.154 con datos de temperatura. Las fuentes de los datos observados suministrados son diversas: Agencia Estatal de Meteorología (AEMET), Confederación Hidrográfica del Guadalquivir (CHG), Cuenca Mediterránea Andaluza (CMA), Servicio de Calidad Ambiental (SIVA), Red de Estaciones Automáticas y Remotas de Meteorología (EARM), Servicio de Evaluación de Recursos Naturales (SERN), Red de Información Agroclimática (RIA), Red de Alerta e Información Fitosanitaria (RAIF), Centro Astronómico Hispano-Alemán (CAHA), Red de Centros Turísticos de Sierra Nevada (CETURSA), Red del Proyecto Guadalfeo (PG) y la Red de Seguimiento del Cambio Global en parques nacionales (RCG).

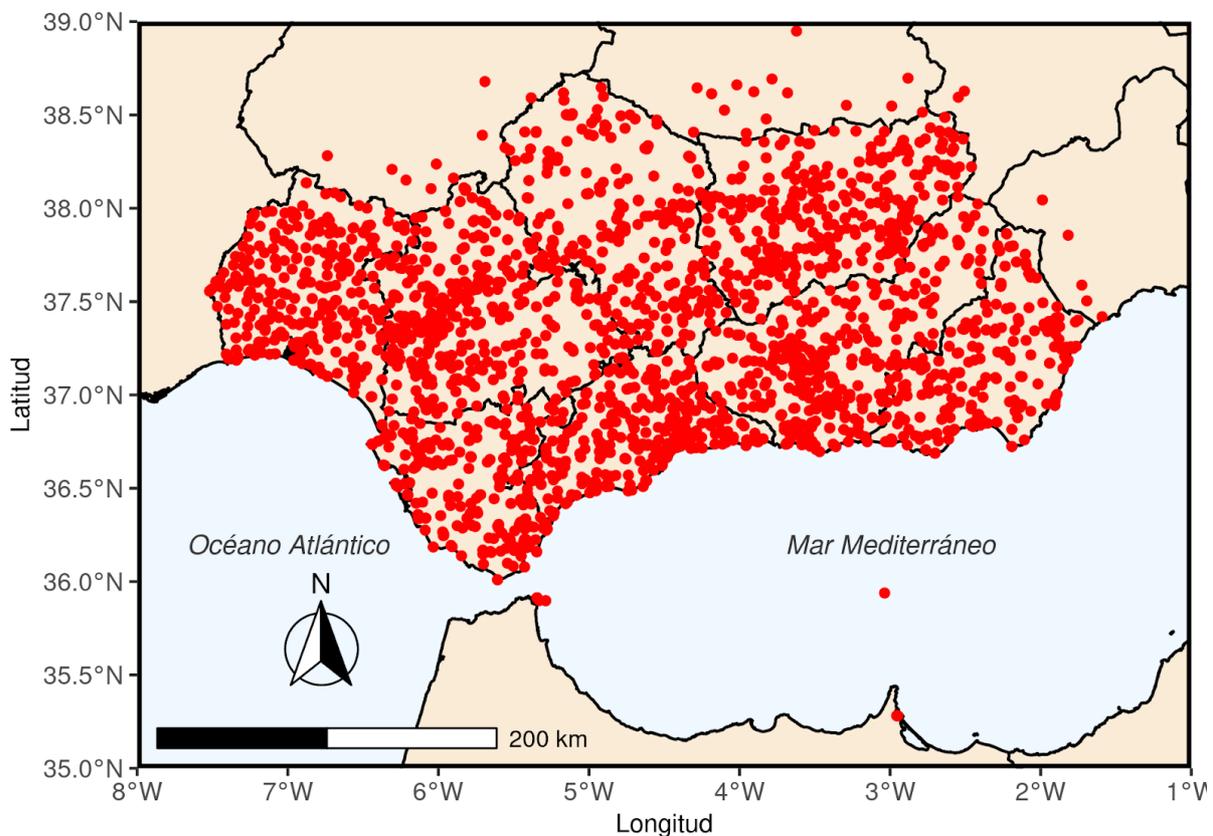


Figura 1.1. Mapa de los observatorios inicialmente suministrados (tanto de precipitación como de temperatura) para el estudio.

Una vez recibidos los datos se procedió a analizar el número de datos diarios de los observatorios, poniendo especial atención a los datos en el periodo 1979-2020, que son los utilizados para realizar la regionalización.

Como se puede observar en la Figura 1.2, hay del orden de 900 observatorios de precipitación y alrededor de 400 observatorios de temperatura que tienen una serie de menos de 2500 datos diarios. La mayor parte de estos observatorios no se podrán utilizar para realizar la regionalización, ya que para que el proceso sea óptimo se necesitan un mínimo de 2000 valores diarios.

También se observa que hay unos 150 observatorios de precipitación y unos 50 de temperatura cuya serie 1979-2020 está prácticamente completa, con unos 15.000 datos diarios disponibles.

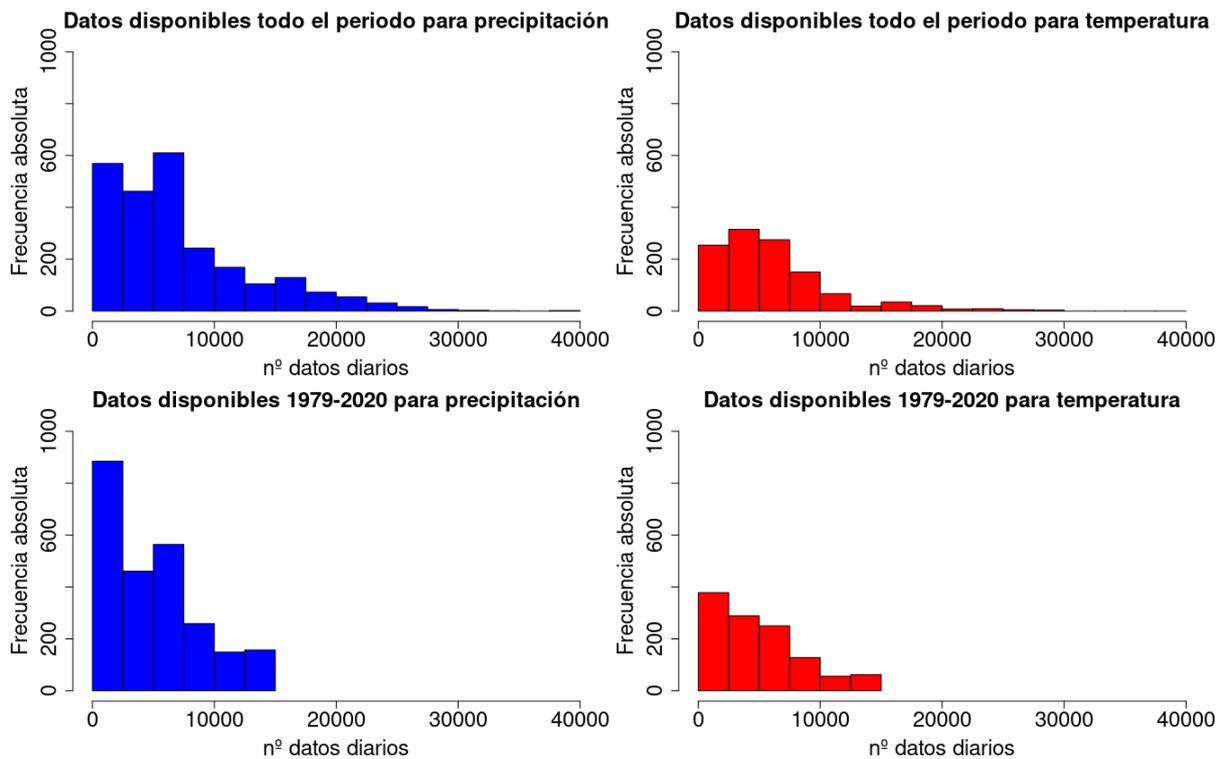


Figura 1.2. Histogramas sobre el número de datos diarios disponibles para las variables de precipitación (columna izquierda, en color azul) y temperatura (columna derecha, en color rojo) para todo el periodo disponible (fila superior) y para el periodo de interés 1979-2020 (fila inferior).

Siguiendo la línea del análisis anterior, se representa el número de estaciones con datos diarios disponibles de más de X años para ambas variables y los mismos períodos, representados en la Figura 1.3. Este análisis corrobora el umbral mínimo de valores diarios necesarios con el fin de emplear el máximo número de datos observados posible, pues cuantos más observatorios se tengan en cuenta, mejor será el resultado de la interpolación. Desde el mismo punto de vista, resulta interesante analizar la cantidad de datos diarios disponibles para cada año, mostrados en la Figura 1.4. El resultado obtenido es favorable, pues los años que poseen la mayor parte de los observatorios corresponden aproximadamente con años del periodo de interés 1979-2020 tanto para temperatura como para precipitación.

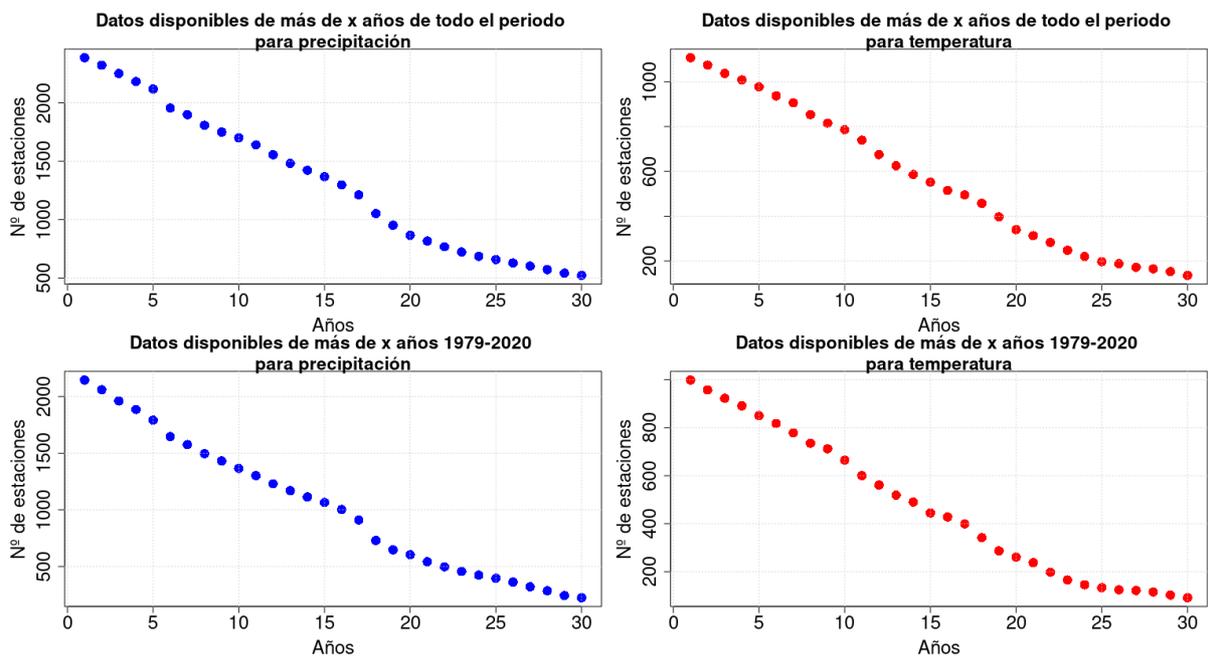


Figura 1.3. Número de estaciones frente a los datos disponibles de más de X años (eje x) para las variables de precipitación (columna izquierda, en color azul) y temperatura (columna derecha, en color rojo) para todo el periodo disponible (fila superior) y para el periodo de interés 1979-2020 (fila inferior).

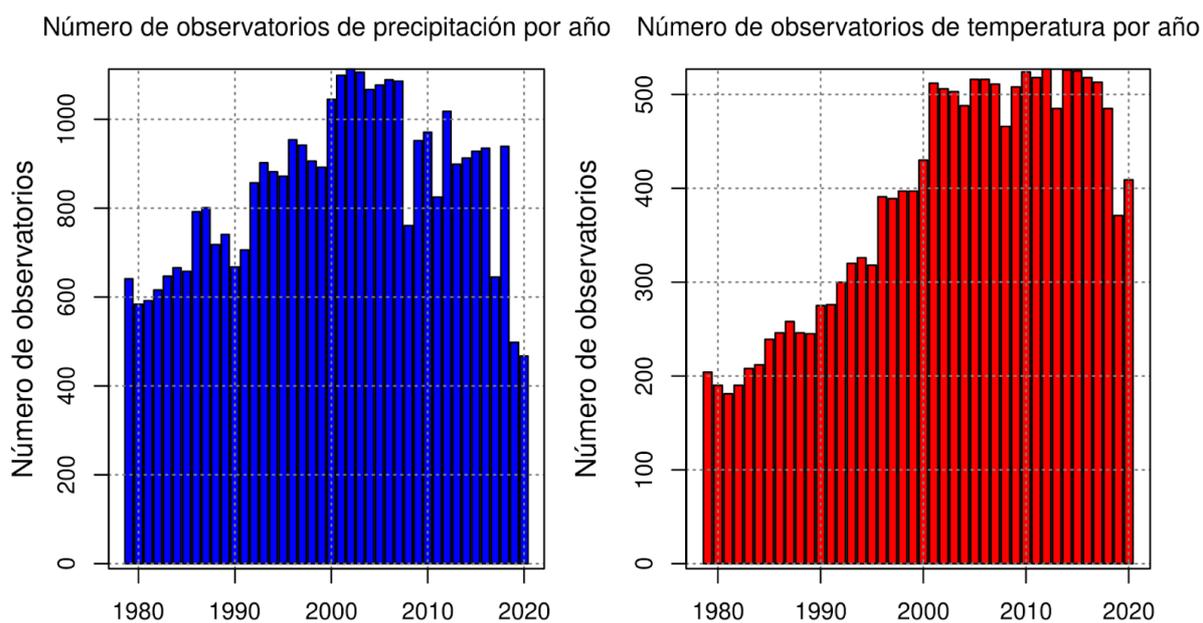


Figura 1.4. Histogramas sobre el número de observatorios de precipitación (columna izquierda, en color azul) y temperatura (columna derecha, en color rojo) disponibles para cada año (eje x) del periodo de interés 1979-2020.

Una vez analizados los datos anteriores, se ha realizado un filtro para eliminar de la lista los observatorios, tanto de temperatura como de precipitación, que no cumplen con la condición indispensable de no tener más de 2000 datos diarios disponibles. La tabla 1.1 muestra las cifras de tal proceso de cribado.

Tabla 1.1. Cifras del cribado de observatorios según el número de datos diarios de los que disponen.

	Número de observatorios con más de 2000 datos	Número de observatorios eliminados
Precipitación	1767	696
Temperatura	827	327

Una vez realizado este filtro, la distribución de observatorios disponibles para realizar el estudio, tanto de temperatura como de precipitación, se muestra en las Figuras 1.5 y 1.6.

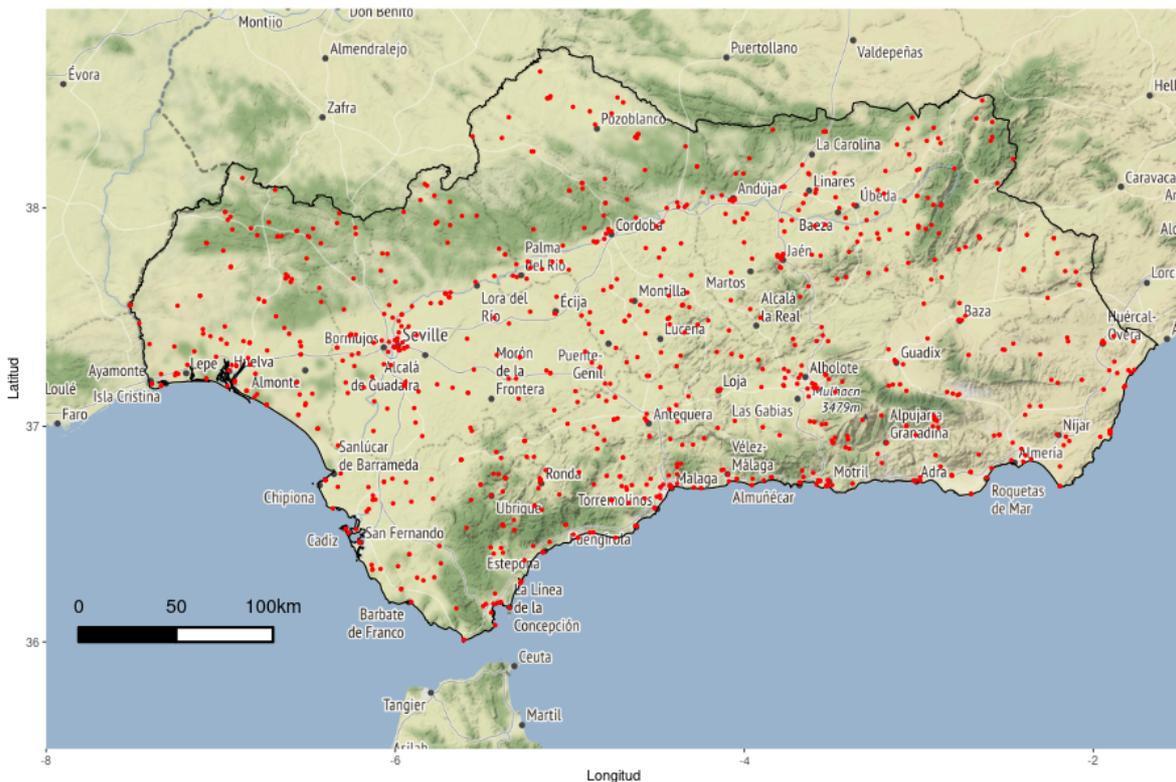


Figura 1. 5. Mapa indicando la localización (puntos rojos) de los 827 observatorios de temperatura seleccionados con más de 2000 datos diarios disponibles.

2. FUNDAMENTO TEÓRICO

2.1. CONTROL DE CALIDAD

Someter a un control de calidad a una serie meteorológica de datos observados consiste en desarrollar un conjunto de pruebas sobre tal serie que garantice que los datos son coherentes dentro de la propia serie de estudio.

Es importante recalcar que tales pruebas deben ser diseñadas de manera que sean capaces de presentar diferentes resultados para diferentes series ya que cada serie recoge la climatología local del lugar observado y por lo tanto aunque los criterios teóricos de las pruebas deban ser los mismos para todos los observatorios los rangos de validez deberán ser dependientes de los observatorios. Si por ejemplo nos basáramos en el estudio de la media de una serie (un posible criterio teórico), que una temperatura máxima diaria fuera de 40°C provoca decisiones diferentes si tal dato pertenece a una serie de media de 35°C (en la que sería un valor más que aceptable) o si tal dato pertenece a una serie de media de 20°C (en la que sería un valor a señalar). Nótese, además, y gracias al anterior ejemplo, que *señalar* un posible valor no implica necesariamente su rechazo, sólo la necesidad de estudiar con más detenimiento tal valor para determinar su sentido.

Los dos controles automáticos principales a emplear en un control de calidad son:

1. Coherencia básica. Rechazo directo de valores manifiestamente erróneos, como por ejemplo las precipitaciones negativas.
2. Puntos atípicos u *outliers*. Valores desacomodadamente atípicos dentro de un conjunto dado de datos, es decir, valores que parecen provenir de fuentes de datos diferentes o haber sido generados de una forma distinta al resto de los datos. En este caso, la dificultad teórica de su detección provendrá de la definición que hagamos de “atípicos”. En la práctica, la detección se refiere generalmente a valores de magnitud absoluta desacomodadamente alta.

2.1.1. TEMPERATURA

COHERENCIA BÁSICA

En el caso de la temperatura, la coherencia básica ha atendido a la búsqueda de valores diarios en los que la temperatura máxima fuera menor que la temperatura mínima. Son dos los tipos básicos de situaciones en los que ocurren estos casos:

1. El dato faltante de temperatura no es recogido como faltante (es decir, con el criterio elegido para reseñar que falta un dato, como “ND” o “-9999”, según la fuente), sino que es consignado como si fuera un 0. En tales casos, los datos de temperatura

máxima pasan directamente a ser rechazados. Nótese que previamente, y del estudio del climograma de la serie, debe determinarse que 0°C es un valor rechazable.

2. El dato de temperatura tiene sentido *por sí mismo* pero *no* al ponerlo en comparación con la otra serie diaria de temperatura; en tal caso, *cualquiera* de las dos series podría ser la errónea, lo que obliga a rechazar el dato diario de ambas.

OUTLIERS

Tal y como se dijo antes, el problema de la detección de un dato atípico radica precisamente en la definición teórica de “atípico”. Fijémonos en la figura 2.1.1.1, mostrando un observatorio (ESNACI) usado como ejemplo de estudio. Parecería evidente que un dato que indica una temperatura máxima diaria de 94°C como la observada en este ejemplo es un *outlier* (y no sólo en un observatorio, desde luego, sino a nivel mundial), pero ¿por qué nos parece que es *evidente*? ¿Sería igual de evidente si fuera de 50°C ? ¿Y si fuera de 40°C ? Desde un punto de vista teórico, la manera de discriminar si un dato es atípico es evaluar cuánto se aleja de los valores típicos de nuestra serie: la manera formal de realizar tal prueba consiste en determinar cuántas veces se aleja un cierto dato de la desviación típica de la serie total observada en cuestión.

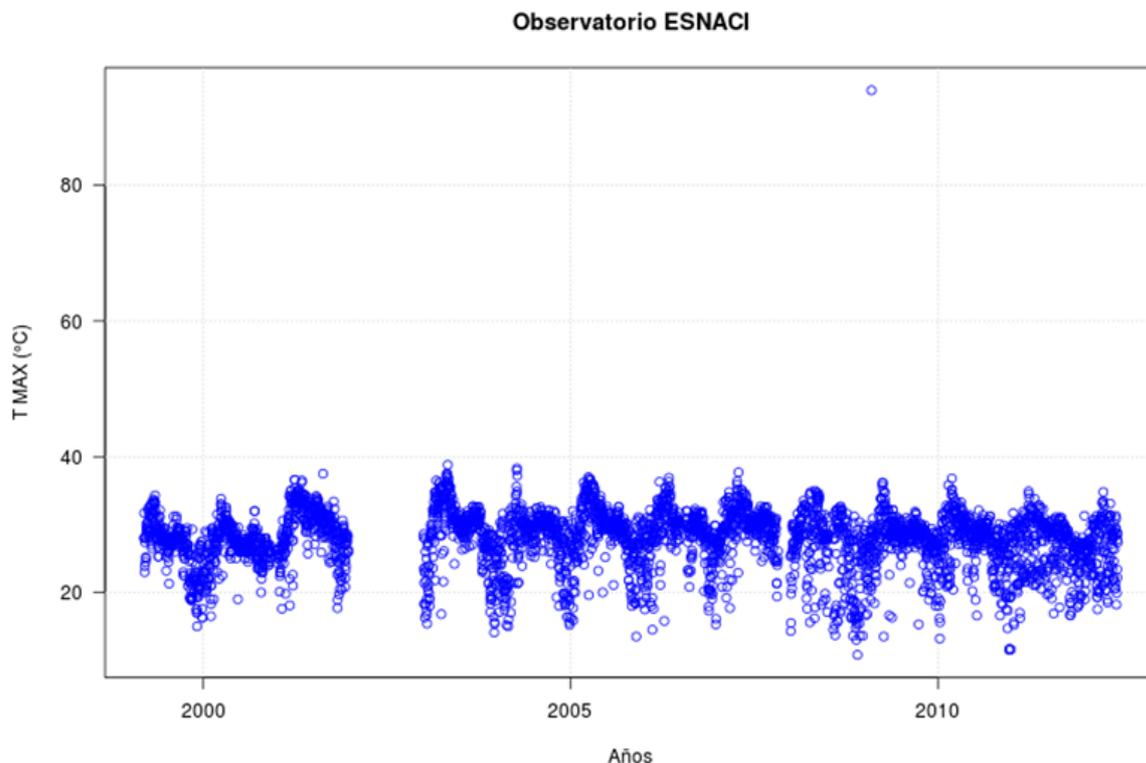


Figura 2.1.1.1. Serie observada diaria de temperatura máxima correspondiente al observatorio ESNACI, mostrando un ejemplo de outlier en una serie temporal.

Por lo tanto, lo que nuestros tests tendrán que determinar será:

1. La desviación típica y la media mensual de cada observatorio.
2. Un “factor desviación” indicando el número de unidades de desviación típica con respecto de la media por encima del cual un cierto valor diario puede ser señalado como atípico.
3. Cálculo de un “valor de desviación” que no debe ser superado según cada distribución mensual [1]:

$$\text{valor desviación} = m \pm \text{factor.desviación} * \text{std} \quad [1]$$

Donde m corresponde con la media del mes del valor a estudiar, std con la desviación estándar del mismo mes y factor.desviación con el número de unidades de desviación estándar establecido en el paso 2.

4. Un estudio directo de los datos señalados como *outliers* por el paso anterior para determinar si son verdaderos y por lo tanto deben ser desechados o si debe reasignarse un nuevo valor umbral y repetirse el estudio.

En el caso de la temperatura, los valores señalados como outliers son muy escasos (algo muy común en el estudio de series de temperatura, que donde suelen presentar problemas es en la etapa de homogeneización, véase más adelante).

2.1.2. PRECIPITACIÓN

COHERENCIA BÁSICA

En el caso de la precipitación, la coherencia básica ha atendido a la búsqueda de valores diarios negativos. Dada la naturaleza evidentemente errónea de estos casos, todos los casos así detectados son inmediatamente marcados como erróneos y desechados.

OUTLIERS

En el caso de la precipitación, un dato anómalo se detecta también mediante la superación de tal dato de un número prefijado de veces de la desviación típica de toda la serie. Sin embargo, y dada la naturaleza de esta variable meteorológica, su detección no implica su eliminación de una manera tan definida como en la temperatura ya que una precipitación extrema es un fenómeno que, aunque inusual, no tiene por qué ser imposible, y necesita por lo tanto de un examen cuidadoso que la contraste con la climatología de la zona. Un ejemplo evidente sería el registro de precipitación diaria en el paso de un huracán (en zonas donde éstos se presenten), que hará que tales valores extremos, aun siendo ciertos, puedan parecer anómalos.

ELIMINACIÓN DE CEROS CONSECUTIVOS

Una característica que se observa en los datos de precipitación es que una serie puede presentar periodos muy largos de meses o años con valores de precipitación cero. En algunas ocasiones, puede haber más de 3 meses sin precipitaciones debido a una sequía extrema; pero en muchos otros casos estos datos son manifiestamente erróneos, pudiendo deberse este error principalmente a que días sin datos son registrados como días de precipitación cero. Hay casos extremos en los que absolutamente toda la serie de precipitación es cero (como por ejemplo en el observatorio real mostrado en la figura 2.1.2.1). Cuando esto ocurre, toda la serie tiene que ser eliminada y no se utilizará en ningún caso.

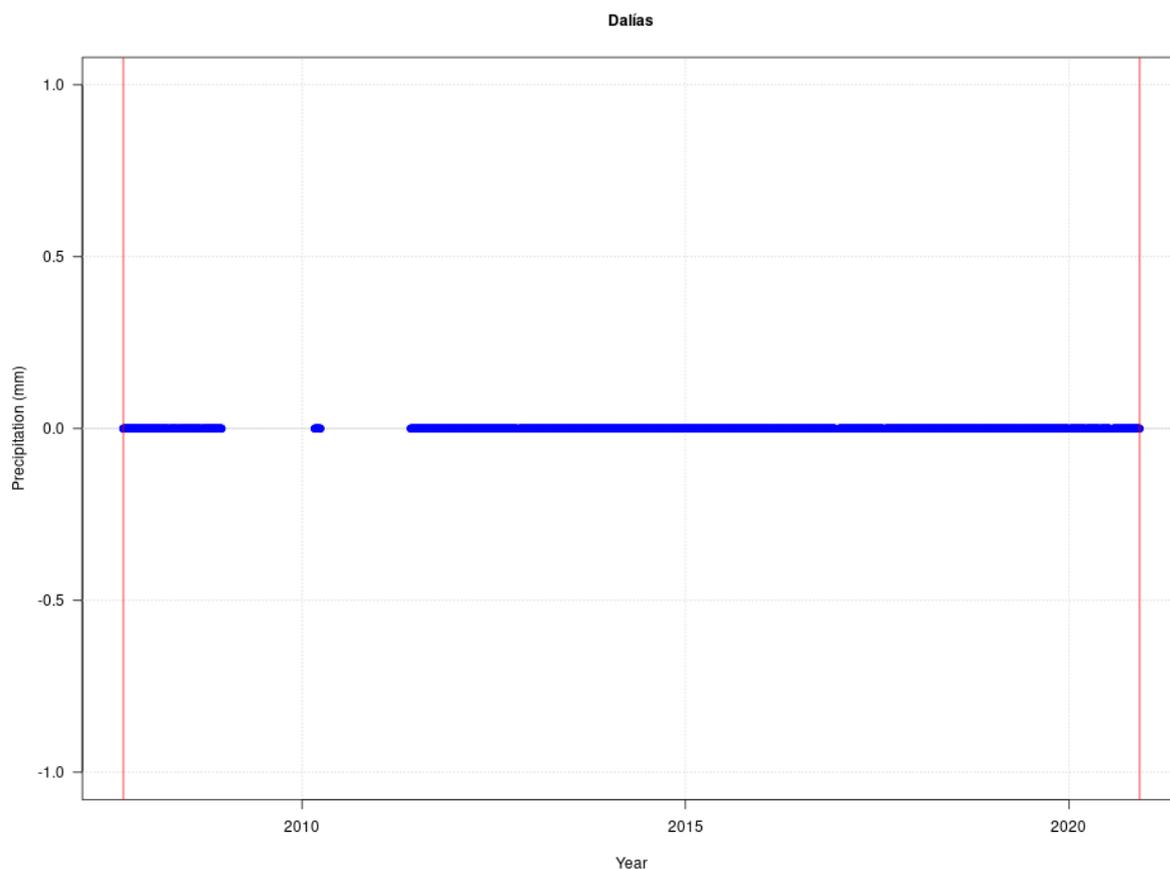


Figura 2.1.2.1. Ejemplo de una serie de precipitación que no ha superado el control de calidad al ser todos sus datos cero.

2.2. HOMOGENEIZACIÓN

La homogeneización de los datos de una serie temporal hace referencia al control de calidad de los datos en tanto que elementos de una serie temporal, es decir, estudia la posible coherencia de los datos exactamente en el orden en el que son presentados. Nótese que los controles anteriores podrían realizarse sobre las mismas series pero desordenadas, pero no nos informan sobre la posible variabilidad temporal de los datos, algo que casi siempre viene ligado cuanto menos a ciclos anuales.

Debe hacerse notar que el proceso de homogeneización de una serie puede presentarse también como parte del proceso general del control de la calidad de los datos de una serie; si se presenta aquí como un punto separado es para subrayar la importancia de tal proceso y de sus resultados.

2.2.1. HOMOGENEIZACIÓN: EJEMPLO TEÓRICO

Fijémonos en la siguiente figura (2.2.1.1.) de temperatura máxima correspondiente a un observatorio real de ejemplo (real pero no perteneciente a este proyecto, sólo lo usamos como ejemplo), con identificador 29015.

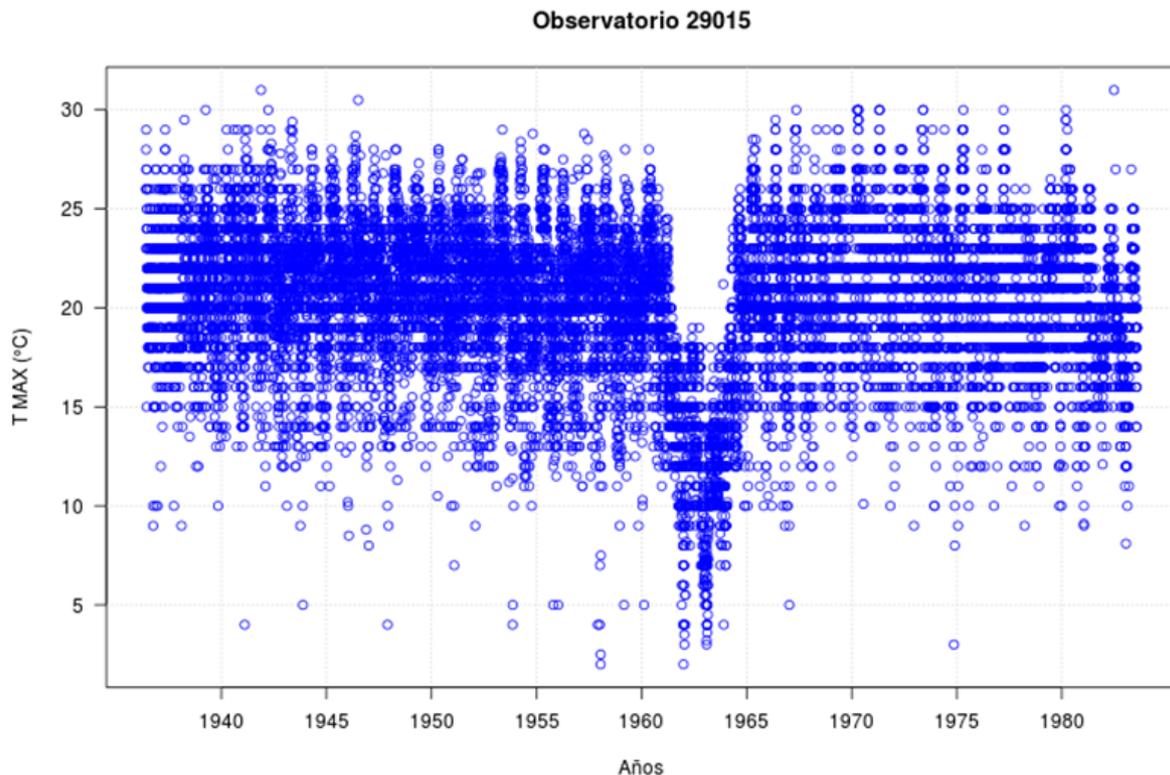


Figura 2.2.1.1. Serie observada diaria de temperatura máxima correspondiente al observatorio ejemplo 29015.

Podemos ver claramente que entre los años 1961 y 1964 este observatorio presenta un comportamiento irregular – irregular si es que lo comparamos con su comportamiento anterior y posterior en el tiempo, pero no por sí mismo ya que los valores que presenta no parecen por sí mismos irracionales. Es por eso por lo que este ejemplo real es ilustrativo de lo que los tests de detección de inhomogeneidades buscan encontrar: fragmentos temporales de la serie que parecen no encajar con el resto de la serie – y de hecho la serie suministrada como ejemplo ha sido encontrada precisamente gracias a estos tests. La dificultad de implementación formal de un test de inhomogeneidad vendrá dada por la definición que hagamos del *parecido* entre un tramo de nuestra serie y el resto de ésta.

La manera de operar del test de homogeneidad que hemos empleado consiste en:

1. Para poder medir cuánto se parece un año a otro, empleamos un test de comparación de distribuciones de Kolmogorov-Smirnov, un test estadístico no-paramétrico (que no presupone distribuciones de la variable a estudiar, por tanto) que nos proporciona un p-valor que podemos emplear como medida del parecido entre dos años. En la figura 2.2.1.2 podemos ver el gráfico “Análisis de control” generado comparando el p-valor de cada año con el año posterior (su logaritmo, en rigor) para el mismo ejemplo anterior (temperatura máxima del observatorio 29015). Valores cercanos a 0 nos indican que dos años tienen una distribución de valores similar entre sí, por lo que podemos concluir que no existe inhomogeneidad entre ellos; a menor valor, más probabilidad de inhomogeneidad entre dos valores consecutivos. Nótese que esta primera parte sólo señala parecidos entre años consecutivos y no es más que una marca previa sobre la posibilidad de existencia de inhomogeneidad.

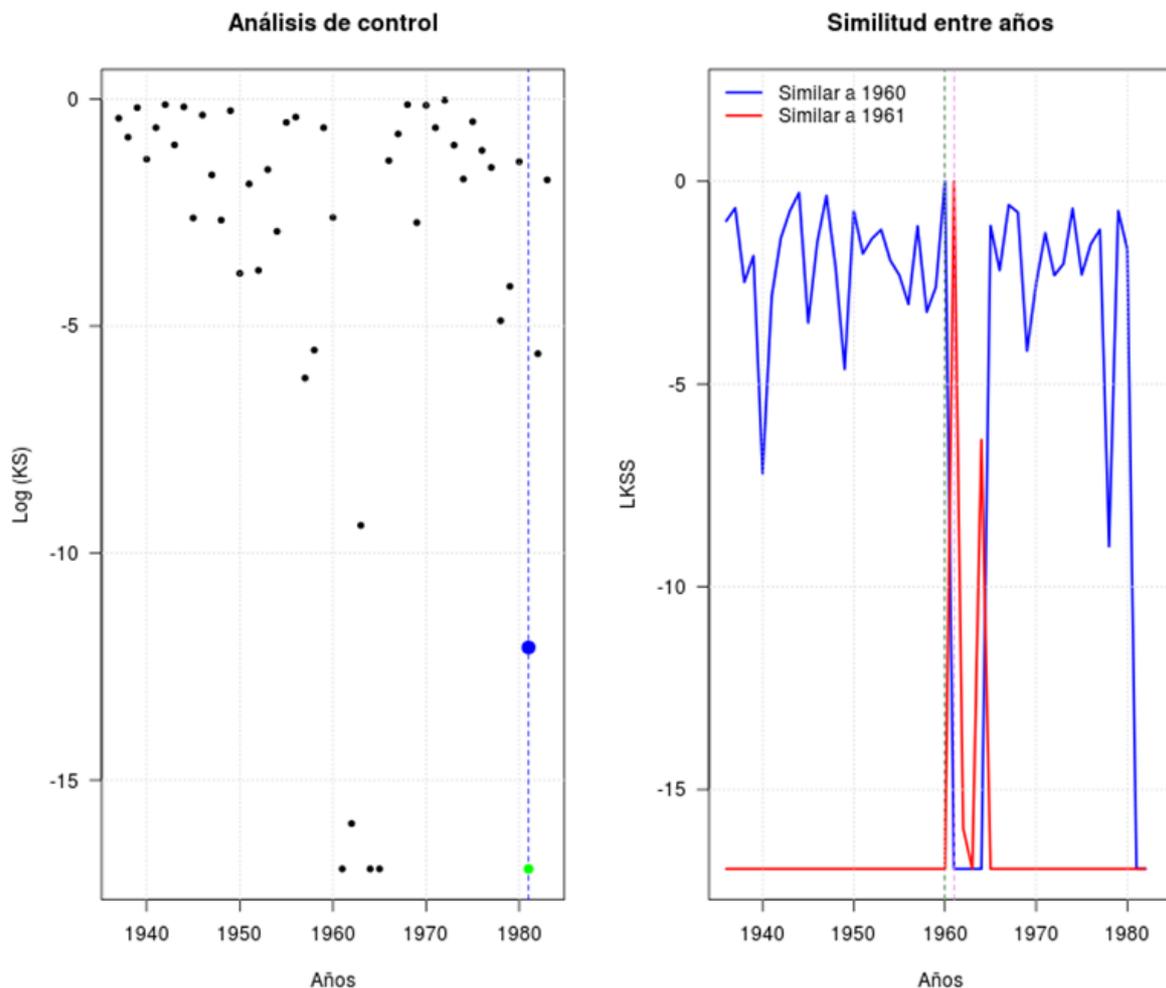


Figura 2.2.1.2. Representación gráfica de la manera de operar del procedimiento empleado de homogeneización de datos diarios. El caso presentado es el correspondiente a la temperatura máxima del observatorio ejemplo 29015, cuyos datos diarios se representan en la Figura 2.2.1.1.

2. Si un cierto año ha sido marcado como posible indicador de inhomogeneidad entonces se le somete a una prueba más genérica (se correspondería con la que aparece marcada como “Similitud entre años” en la figura 2.2.1.2.). Una vez marcados el año de corte y el siguiente (que sabemos de diferente distribución y presuponemos marcan periodos diferentes) averiguamos el p-valor de cada uno de los años de la serie con respecto a esos dos años (se corresponderían con las líneas roja y azul de la gráfica). Si se presenta un salto o una ruptura entre todos esos p-valores en los años que estamos considerando entonces sí que se considera que tenemos una inhomogeneidad verdadera para toda la serie.

Esta manera de trabajar permite señalar en qué años aparece una cierta inhomogeneidad en una serie. Dado que establecer lo pequeño que es un p-valor para así señalar posible inhomogeneidad es una cuestión de criterio, se ha lanzado varias veces el mismo tests con diferentes p-valores de corte (de muy negativos a más cercanos a 0) para así eliminar la subjetividad del criterio, ya que una inhomogeneidad, de ser cierta, debería aparecer en la mayoría de las ejecuciones de los tests.

2.2.2. EJEMPLOS

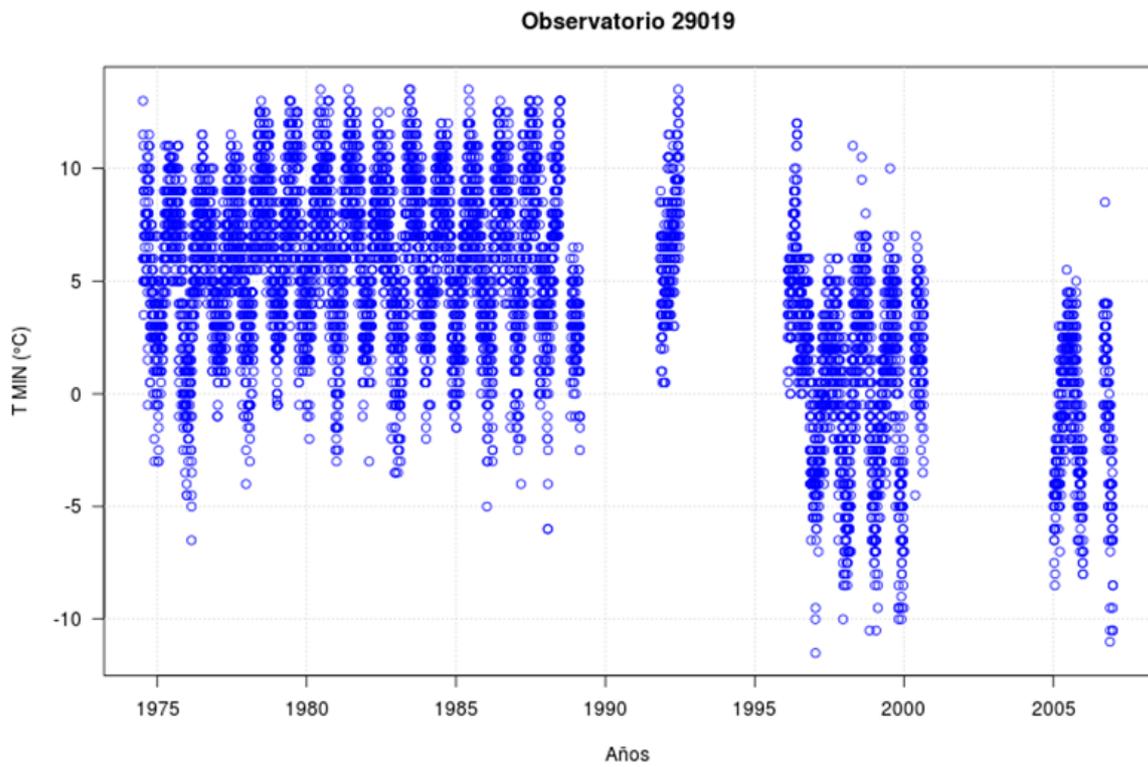
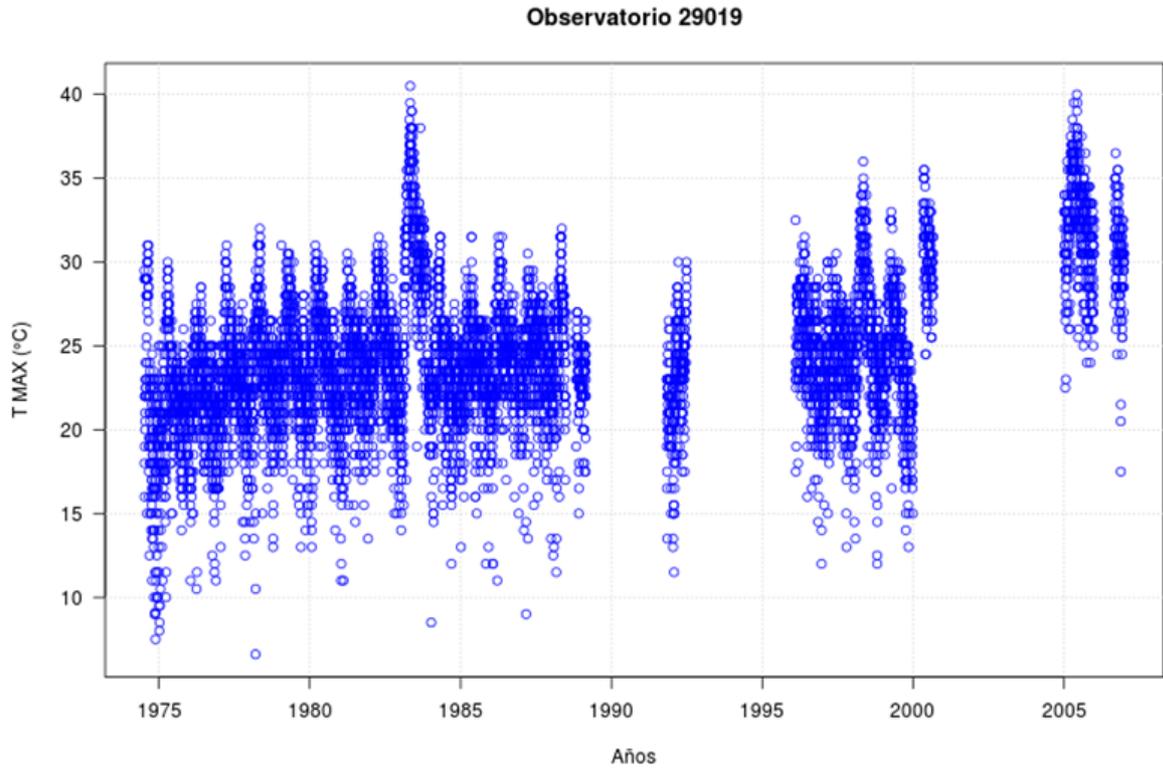


Figura 2.2.2.1. Representación gráfica de las series de temperatura máxima diaria y temperatura mínima diaria del observatorio 29019.

Dada la peculiaridad de un estudio de homogeneización, lo diferentes que pueden ser sus resultados, y lo diferentes que pueden ser los criterios de actuación según los resultados obtenidos, se han incluido una serie de ejemplos ilustrativos con series reales proporcionadas para este estudio que provocaron señales de inhomogeneidad al serles aplicados estos tests:

1. Figura 2.2.2.1. Un mismo observatorio (29019) presenta varias inhomogeneidades tanto en la serie de temperatura máxima como en la de mínima. Por una parte, la serie de temperatura máxima presenta un fuerte (y evidente) salto en 1983. Pero los tests también señalan que los valores a partir de 1996 no siguen la misma pauta que los anteriores, algo aproximadamente visible (los valores parecen tener un sesgo añadido). Parecería necesario eliminar el año 1983 y darlo como erróneo y eliminar los años posteriores a 1996 – o los anteriores. Pero si nos fijamos en la temperatura mínima diaria de tal observatorio observamos una fuerte discrepancia entre los años anteriores a 1996 y los posteriores: de nuevo habría que elegir cuál es el periodo “correcto” y cuál el “incorrecto” y eliminarlo. Dado que este observatorio de temperatura presenta estas fuertes inhomogeneidades en las dos variables medidas y que la inspección directa no permite asegurar cuál es el periodo correcto, hemos optado por eliminarlo de nuestro estudio.
2. Figura 2.2.2.2. La variable temperatura máxima diaria del observatorio ESNACI, de nuevo, pero una vez eliminado el *outlier* detectado en el apartado anterior. Los tests nos indican inhomogeneidad entre los datos correspondientes a los años del periodo anterior al año 2000 y los posteriores; una inspección visual nos indica que, efectivamente, los dos primeros años no parecen mostrar una distribución similar a todos los posteriores. En este caso, la serie no es eliminada sino corregida: se eliminan los datos de los dos primeros años (se consideran como erróneos) y se consideran como correctos todos los datos restantes.
3. Figura 2.2.2.3. Un ejemplo de detección de inhomogeneidad en la variable de precipitación diaria: los tests señalan, para este observatorio (21096), una inhomogeneidad entre los datos pertenecientes al periodo anterior al año 1950 y los posteriores. Dada la variabilidad de la precipitación diaria, el estudio gráfico directo de las series de precipitación rara vez ofrece conclusiones visibles; en este caso, sí que el primer núcleo cercano al año 1950 parece tener una distribución diferente del resto del años y se ha optado por eliminar esos años pero el poder tomar esta decisión es poco común – hay que comprender que dos años seguidos en el mismo observatorio pueden tener regímenes hídricos muy diferentes, provocar una alarma automática de inhomogeneidad y ser, sin embargo, ciertos (por ejemplo: un cambio hacia un periodo de sequía), pero también es cierto que una alarma de inhomogeneidad siempre quiere llamarnos la atención sobre la diferencia entre la distribución de los datos de dos

periodos, por lo tanto debemos marcar los observatorios con tal efecto como de fiabilidad media.

4. Figura 2.2.2.4. El observatorio que usamos como ejemplo para abrir este apartado, una vez corregida su serie de temperatura máxima. Obsérvese cómo al eliminar los años problemáticos la serie, aunque ahora incompleta, presenta homogeneidad en todo el periodo dado como correcto, algo que es apreciable gráficamente y que los tests de homogeneidad también muestran.

Es importante reseñar, como conclusión final, que no existe una metodología automática que garantice siempre la corrección de la acción a ejecutar sobre la serie señalada (eliminación o corrección); pero siendo como son necesarios cuando se parte de grupos grandes de observatorios (como es nuestro caso) siempre necesitan de una evaluación manual posterior para asegurar la corrección de las series.

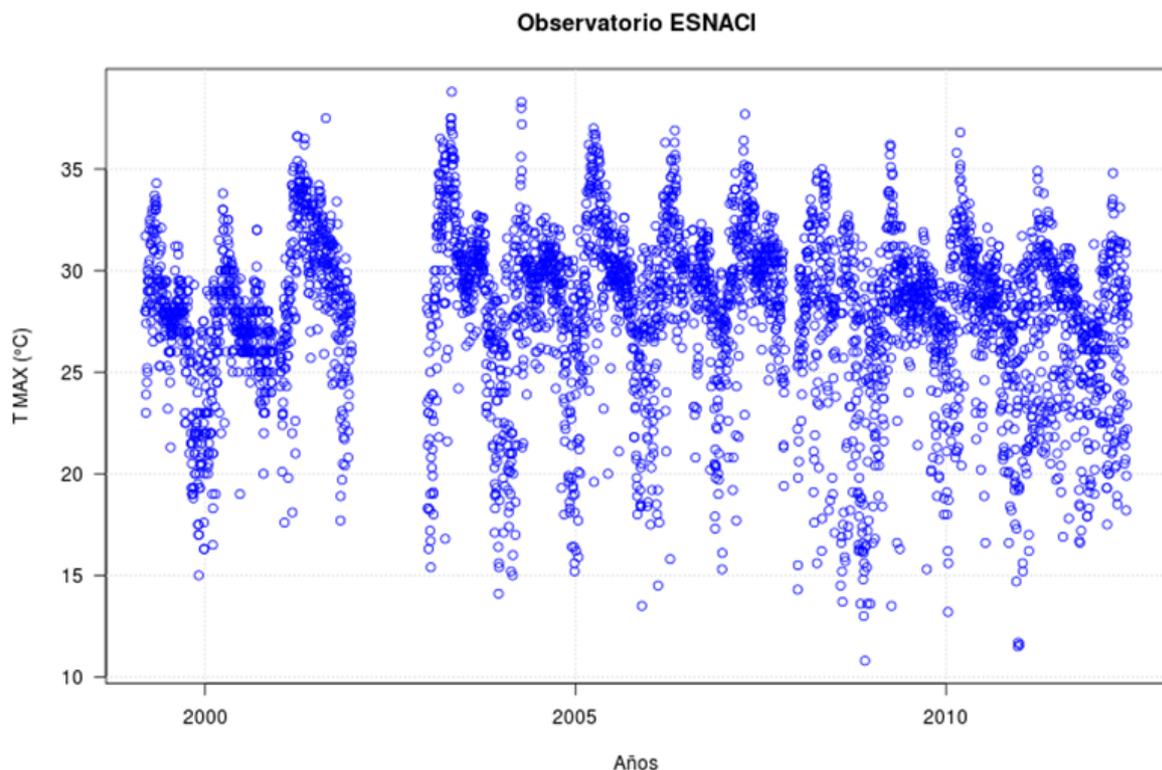


Figura 2.2.2.2. Representación gráfica de la serie de temperatura máxima diaria del observatorio ESNACI (obsérvese que se ha eliminado el *outlier* señalado en la figura Figura 2.1.1.1).

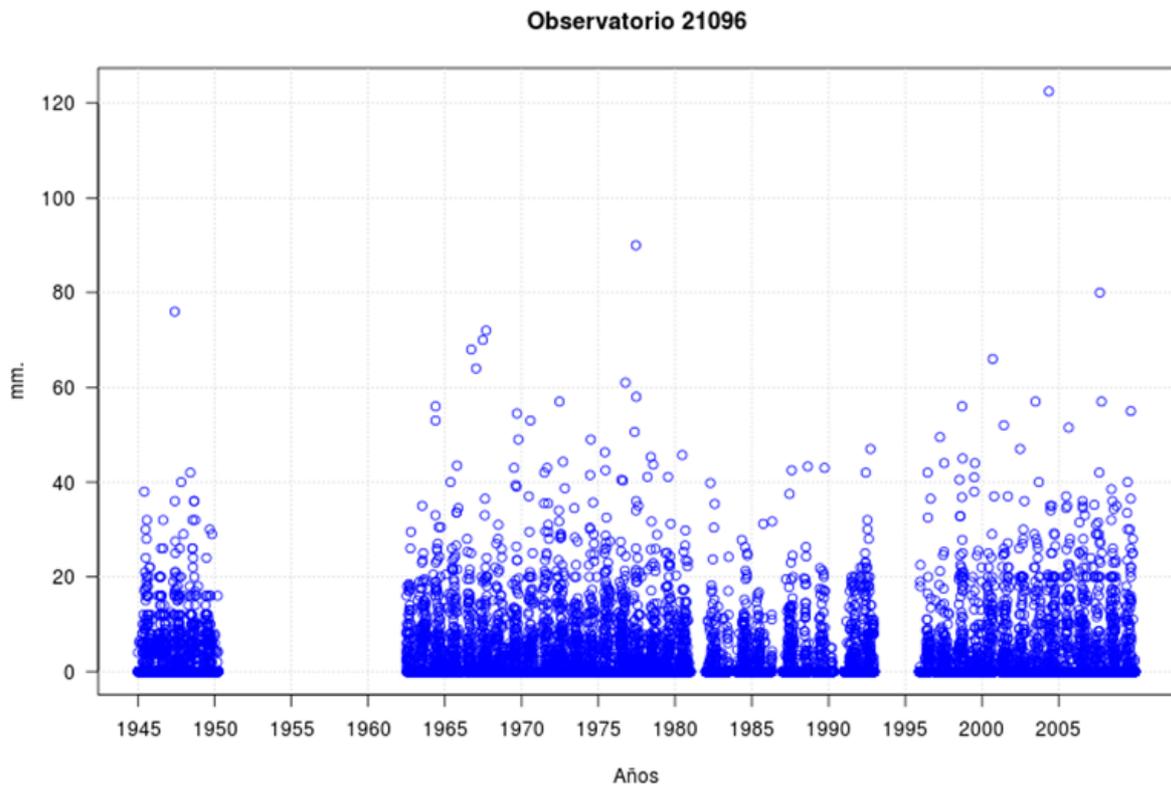


Figura 2.2.2.3. Representación gráfica de la serie de precipitación diaria del observatorio 21096.

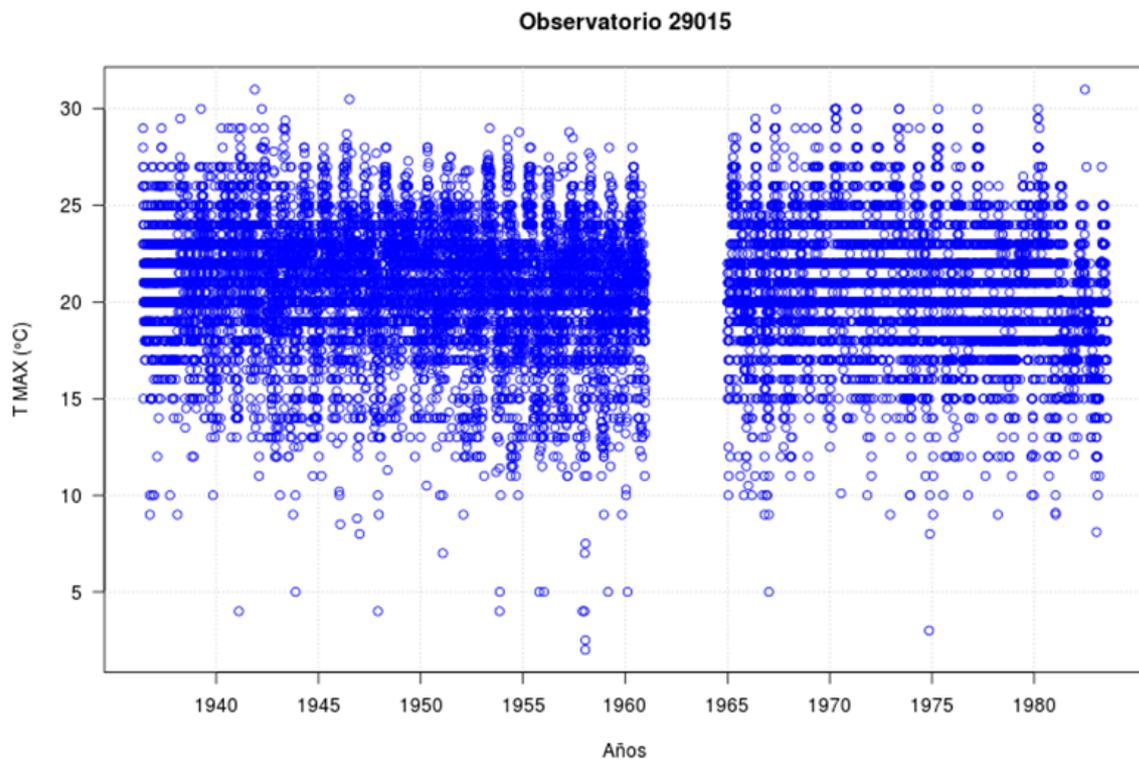


Figura 2.2.2.4. Representación gráfica de la serie de temperatura máxima diaria del observatorio 29015 una vez eliminados los años determinados como incorrectos (compárese con los datos originales representados en la figura 2.2.1.1.).

3. TEMPERATURA

3.1. CONTROL DE CALIDAD

Para garantizar la coherencia de los datos observados procedentes de cada estación se realiza un control de calidad desde los dos puntos de vista introducidos anteriormente. Este proceso se realiza de forma automatizada dado la cantidad de observatorios del caso de estudio.

3.1.1. COHERENCIA BÁSICA

Consiste en la detección y sustitución por valores NA (sin dato) de medidas diarias en las que el valor de temperatura mínima diaria sea superior al valor de temperatura máxima diaria, pues al no conocer el motivo de este error es preferible que sean eliminados antes que corregidos erróneamente. En este sentido, 41 de los observatorios fueron corregidos.

3.1.2. OUTLIERS. VALORES EXTREMOS.

Detección y eliminación de valores extremos que no constan haber sido superados por registros oficiales. Este criterio ha sido establecido considerando como orientativos los valores extremos de las Efemérides de observatorios de AEMET [http://www.aemet.es/es/serviciosclimaticos/datosclimatologicos/efemerides_extremos?w=0&k=and&datos=img]. De los 16 observatorios disponibles en el registro para el territorio andaluz, se obtiene que el valor más alto de temperatura fue alcanzado en el observatorio del aeropuerto de Córdoba alcanzando los 46.9°C en 2017, mientras que el valor más bajo de temperatura fue registrado en el observatorio del aeropuerto de Granada, alcanzando los -14.2°C en 1987. Estos valores han sido orientativos a la hora de establecer los límites de valor máximo de temperatura máxima y valor mínimo de temperatura mínima, pues posteriormente han sido ligeramente modificados a la hora de establecer el intervalo de valores aptos teniendo en cuenta la variedad de observatorios proporcionados por la Junta de Andalucía para el análisis, pues se tienen, por ejemplo, observatorios situados en la Sierra de Cazorla o Sierra Nevada que han podido registrar valores aún menores que -14.2°C; asimismo, se modifica el límite inferior de las temperaturas máximas ampliando el intervalo según la altitud del observatorio en cuestión. Para establecer los límites máximo mínimo y mínimo máximo se consultan fuentes de datos diferentes:

[<http://www.tiempoensevilla.es/wxcurecords.php>]

[https://www.meteoblue.com/es/tiempo/historyclimate/climatemodelled/estaci\%c3\%b3n-de-esqu\%c3\%ad-de-sierra-nevada_esp\%c3\%b1a_7602397]

Así, los umbrales establecidos son los siguientes:

- Altitud inferior a 370 m:
 - Temperatura mínima: (valor máximo: 30°C, valor mínimo -20°C)

- Temperatura máxima: (valor máximo: 48°C, valor mínimo: 0°C).
- Altitud 370 m - 420 m:
 - Temperatura mínima: (valor máximo: 30°C, valor mínimo -20°C)
 - Temperatura máxima: (valor máximo: 48°C, valor mínimo: -1°C)
- Altitud 420 m - 900 m:
 - Temperatura mínima: (valor máximo: 30°C, valor mínimo -20°C)
 - Temperatura máxima: (valor máximo: 48°C, valor mínimo: -5°C)
- Altitud superior a 900 m:
 - Temperatura mínima: (valor máximo: 30°C, valor mínimo -20°C)
 - Temperatura máxima: (valor máximo: 48°C, valor mínimo: -15°C)

Altitud (m)	Variable	Valor mínimo	Valor máximo
< 370	Temperatura máxima	0°C	48°C
	Temperatura mínima	-20°C	30°C
370-420	Temperatura máxima	-1°C	48°C
	Temperatura mínima	-20°C	30°C
420-900	Temperatura máxima	-5°C	48°C
	Temperatura mínima	-20°C	30°C
> 900	Temperatura máxima	-15°C	48°C
	Temperatura mínima	-20°C	30°C

De acuerdo con estos umbrales, 446 observatorios fueron corregidos (por ejemplo Figuras 3.1.2.1 y 3.1.2.2. En esta última figura se encuentra un problema recurrente en los observatorios RAIF, en el que en lugar de indicar NA o -9999 cuando no hay dato, se indica +99999).

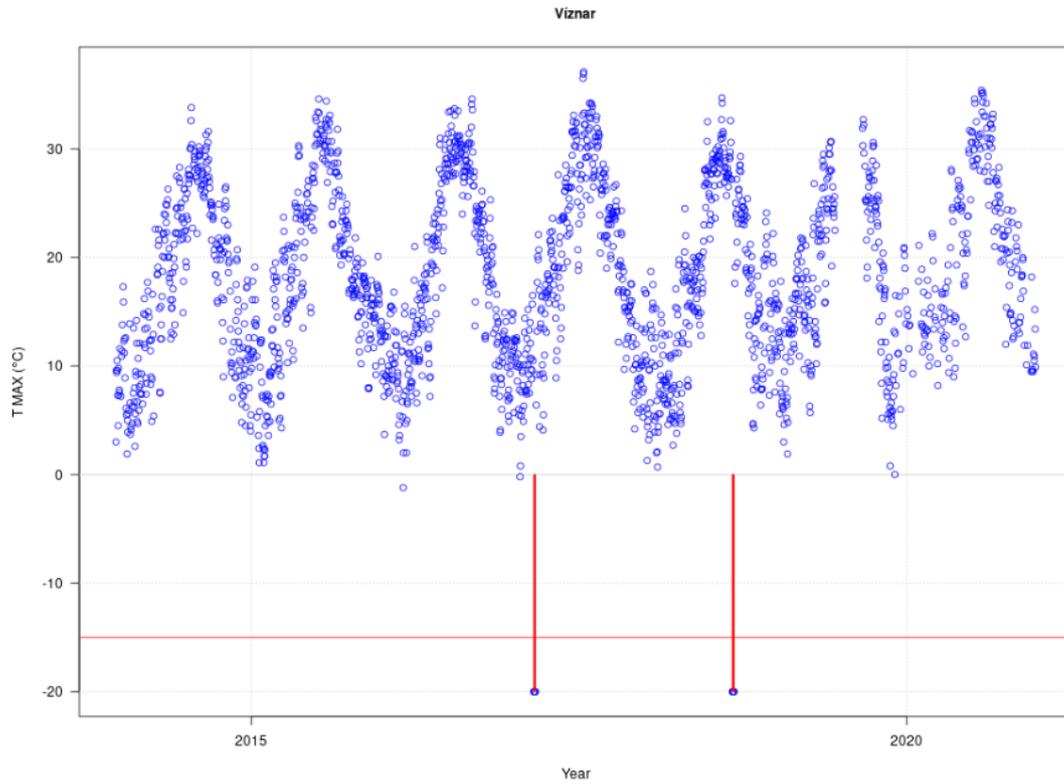


Figura 3.1.2.1. Representación gráfica de la serie de temperatura máxima diaria del observatorio situado en Viznar. Ejemplo de observatorio en el que se han corregido outliers por valores extremos.

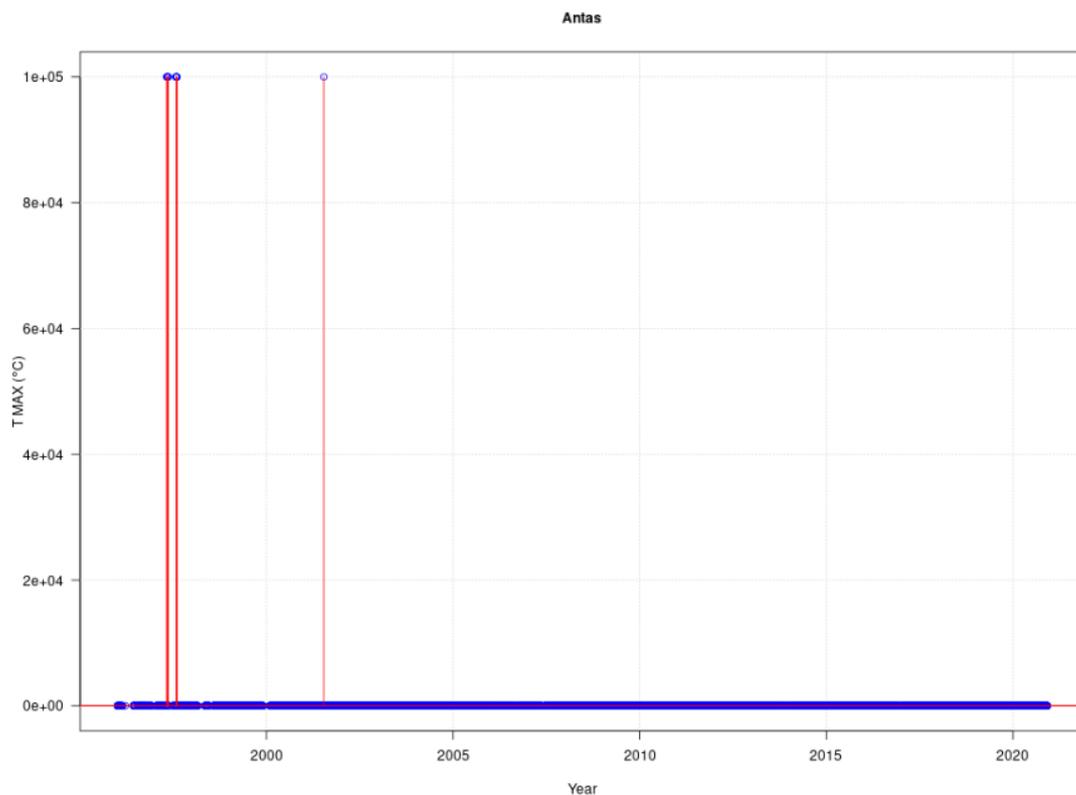


Figura 3.1.2.2. Representación gráfica de la serie de temperatura máxima diaria del observatorio situado en Antas. Ejemplo de observatorio en el que se han corregido outliers por valores extremos.

3.1.3. OUTLIERS. LÍMITE DISTRIBUCIÓN MENSUAL.

Detección y eliminación de valores superiores al “valor desviación” definido por la ecuación [1] (ver apartado 2.1.1). Para la variable de temperatura se considera un “factor desviación” de 7 unidades. En este sentido, 98 de los observatorios fueron corregidos (por ejemplo Figura 3.1.3.1.).

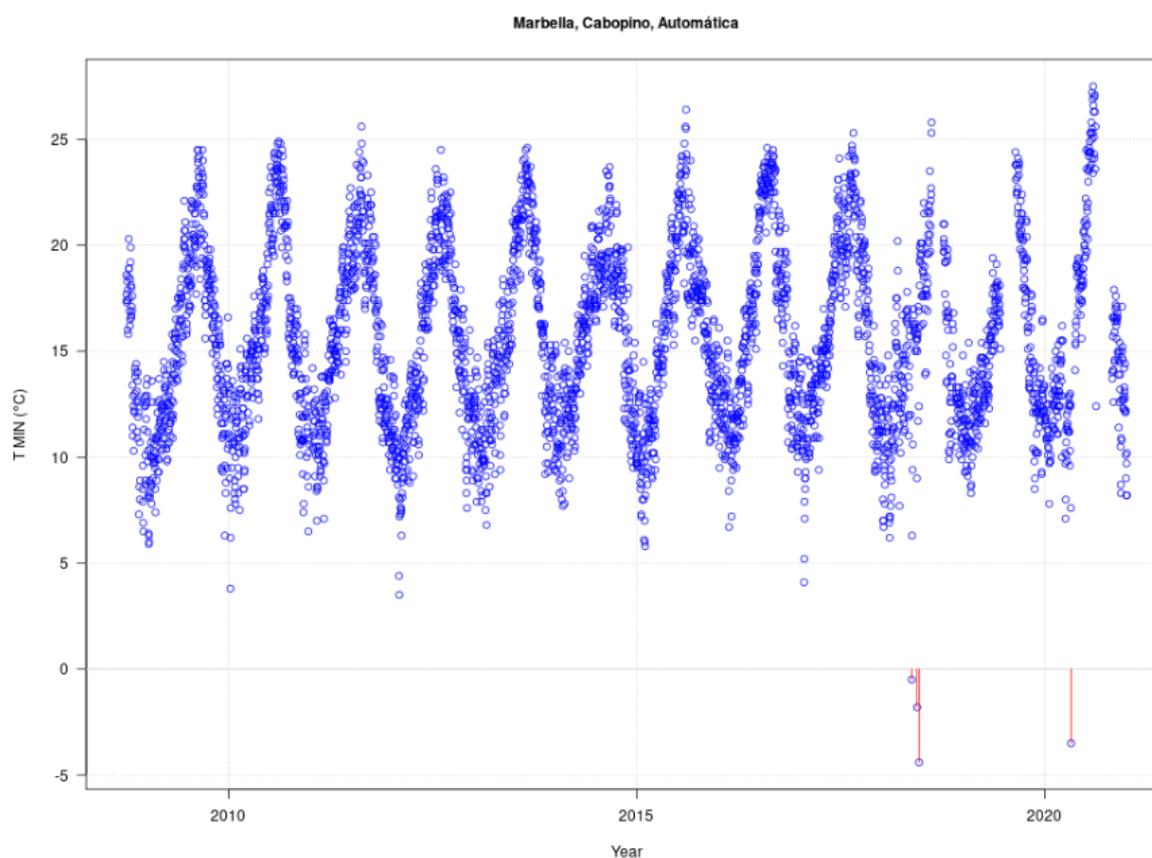


Figura 3.1.3.1. Representación gráfica de la serie de temperatura mínima diaria del observatorio situado en Marbella, Cabopino. Ejemplo de observatorio en el que se han corregido outliers por límite de distribución mensual.

3.2. HOMOGENEIZACIÓN

Tras haber realizado el control de calidad, los valores restantes son analizados desde el punto de vista de la homogeneidad, en el que se analizan y corrigen si es necesario la temperatura máxima diaria y la temperatura mínima diaria por separado. Una vez detectados los saltos de homogeneidad tras aplicar el test de homogeneidad de Kolmogorov-Smirnov y con el fin de automatizar un proceso de homogeneización, se consideran las siguientes situaciones y acciones:

- Si hay salto de homogeneidad anterior a 1979, la serie no es modificada.
- Si hay salto de homogeneidad posterior a 1979, se diferencian dos subseries: una desde el final de la serie hasta el salto más reciente y otra desde el salto más reciente hasta el salto anterior o, en caso de no haberlo, hasta el inicio de la serie. Entonces se comparan las medias de las dos subseries:
 - Si la diferencia de medias es inferior a 2°C, la serie no es modificada.
 - Si la diferencia de medias es superior a 2°C y la longitud de datos de la subserie más reciente es igual o inferior a 2500 datos diarios, la serie no es modificada.
 - Si la diferencia de medias es superior a 2°C y la longitud de datos de la subserie más reciente es superior a 2500 datos diarios, todo lo anterior a la última subserie es eliminado para ese observatorio.

Si la serie analizada posee más de un salto de homogeneidad posterior a 1979, se repite el mismo proceso hasta completar todas las subseries con singularidad posterior a 1979. Como se puede observar, en el criterio establecido se considera que los datos más recientes son los más realistas.

A continuación se muestran tres ejemplos indicando cómo se ha procedido en el proceso de homogeneización.

En el primer caso (Figura 3.2.1.), el test de homogeneidad aplicado detecta una singularidad en el año 1991 (es decir, posterior a 1979). Las líneas horizontales discontinuas de color negro muestran la media de cada uno de los subperiodos que el test ha encontrado en la serie del observatorio, entre las que se observa una diferencia inferior a 2°C. Así, según las indicaciones del proceso no se modificaría la serie mostrada. También se puede observar claramente un problema de resolución muy recurrente en gran parte de los observatorios: desde el año 1976 se observa un cambio importante en la resolución de medición de la temperatura, pasando de una resolución de 0.2°C a 1°C.

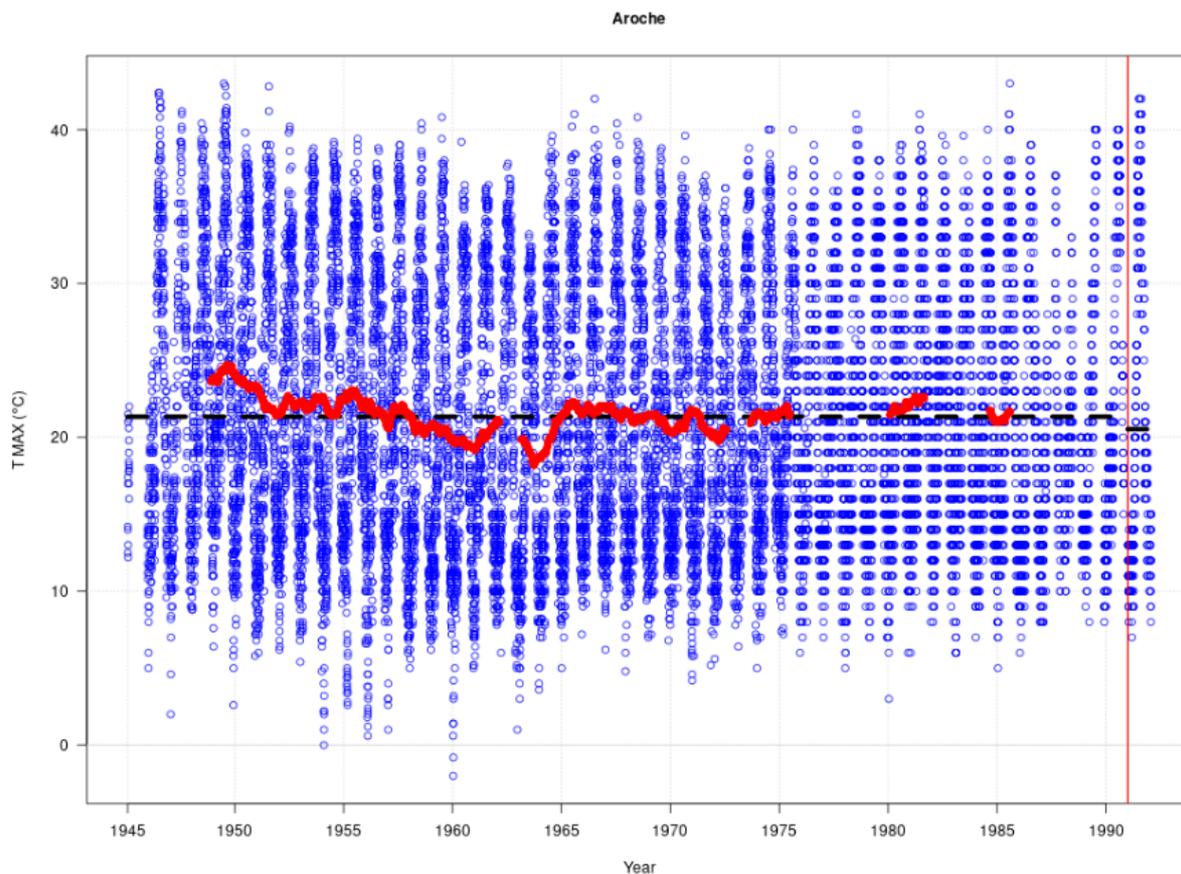


Figura 3.2.1. Representación gráfica de la serie de temperatura máxima diaria del observatorio situado en Aroche. Ejemplo de procedimiento del test de homogeneidad.

En el segundo caso (Figura 3.2.2.), el test de homogeneidad encuentra dos singularidades, una en el año 1971 y otra en el año 1981. Centrando la atención en primer lugar en los dos subperiodos más recientes, se observa que la diferencia de medias es superior a 2°C. Además, como la longitud de datos de la subserie más reciente es superior a 2500 datos diarios, todo lo anterior a la última subserie es eliminado para este observatorio. Cabe destacar de nuevo el problema en la resolución de medición, que en este caso sufre más de un cambio.

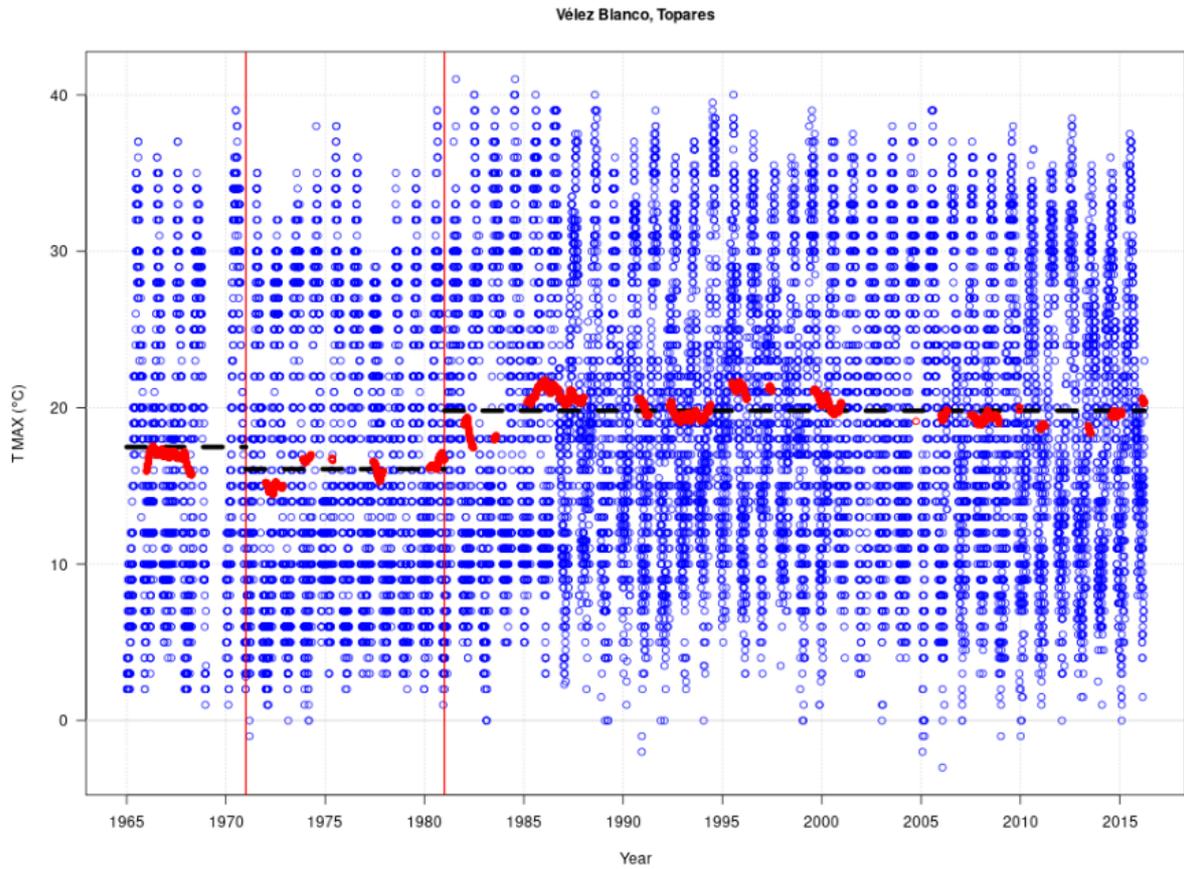


Figura 3.2.2. Representación gráfica de la serie de temperatura máxima diaria del observatorio situado en Vélez-Blanco, Topares. Ejemplo de procedimiento del test de homogeneidad.

En el tercer caso (Figura 3.2.3.), el test de homogeneidad encuentra una singularidad en el año 1976. Como se trata de un salto en la homogeneidad anterior a 1979, la serie no es modificada.

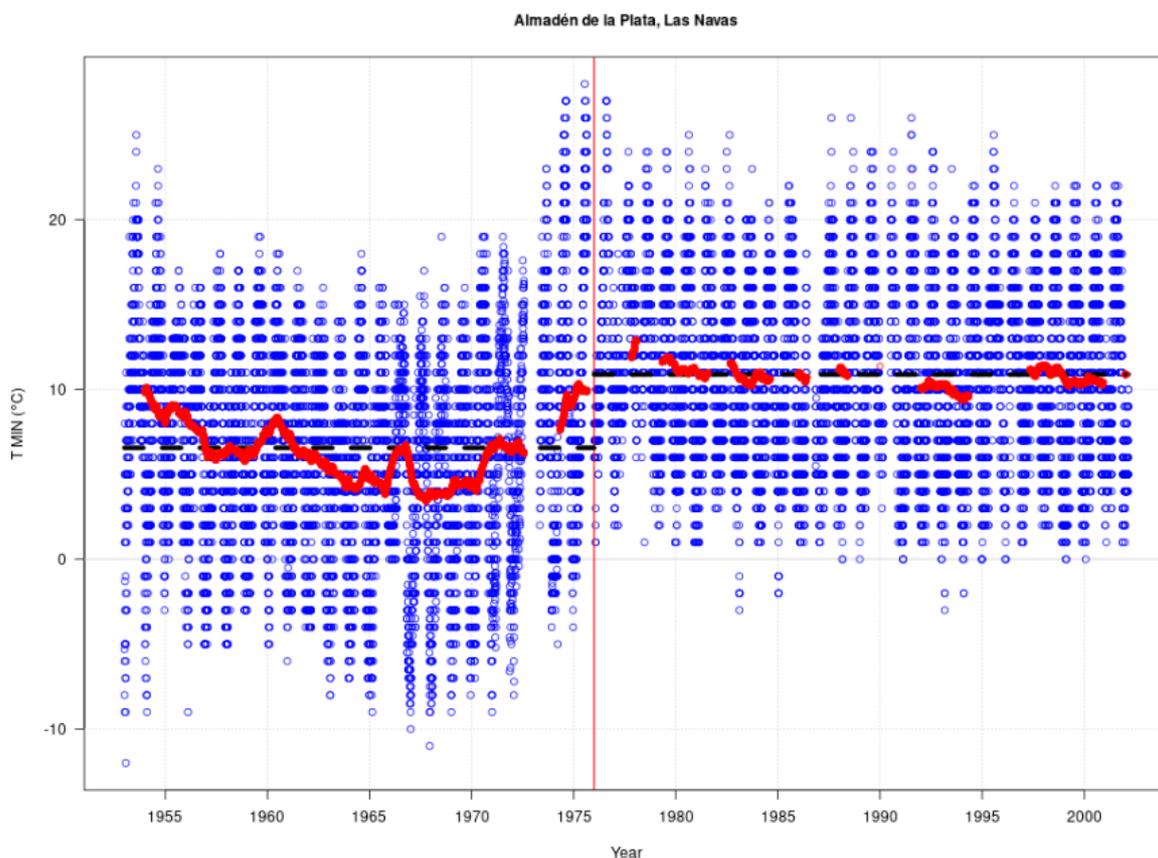


Figura 3.2.3. Representación gráfica de la serie de temperatura mínima diaria del observatorio situado en Almadén de la Playa, Las Navas. Ejemplo de procedimiento del test de homogeneidad.

Por otro lado, tras haber aplicado el test de homogeneidad se decide eliminar 8 observatorios que proporcionan datos muy dispares y poco fiables (por ejemplo Figuras 3.2.1. y 3.2.2.); también se decide modificar manualmente irregularidades destacables que se encuentran en los valores de 14 observatorios y que, a su vez, son muy difíciles de eliminar de forma automatizada (por ejemplo Figura 3.2.3). Por último, se detectan 30 observatorios duplicados (situados en la misma localización que otros) que son eliminados atendiendo a criterios de fuente de información, periodo disponible de observaciones y continuidad de actividad. De esta manera, de los 827 observatorios de temperatura pre-seleccionados, hay 789 que han superado los controles y cuyas series podrán ser utilizadas para el estudio.

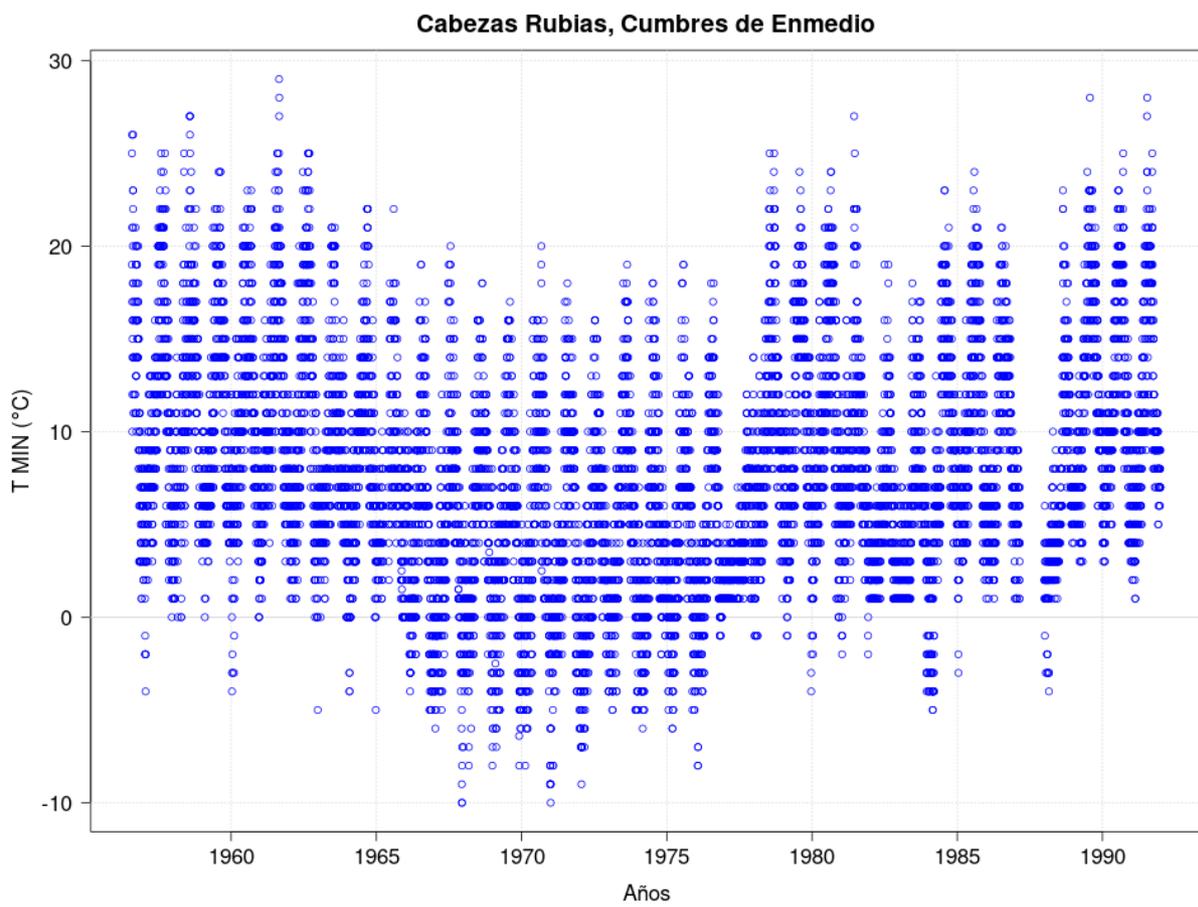


Figura 3.2.4. Representación gráfica de la serie de temperatura mínima diaria del observatorio situado en Cabezas Rubias, Cumbres de Enmedio. Ejemplo de observatorio eliminado.

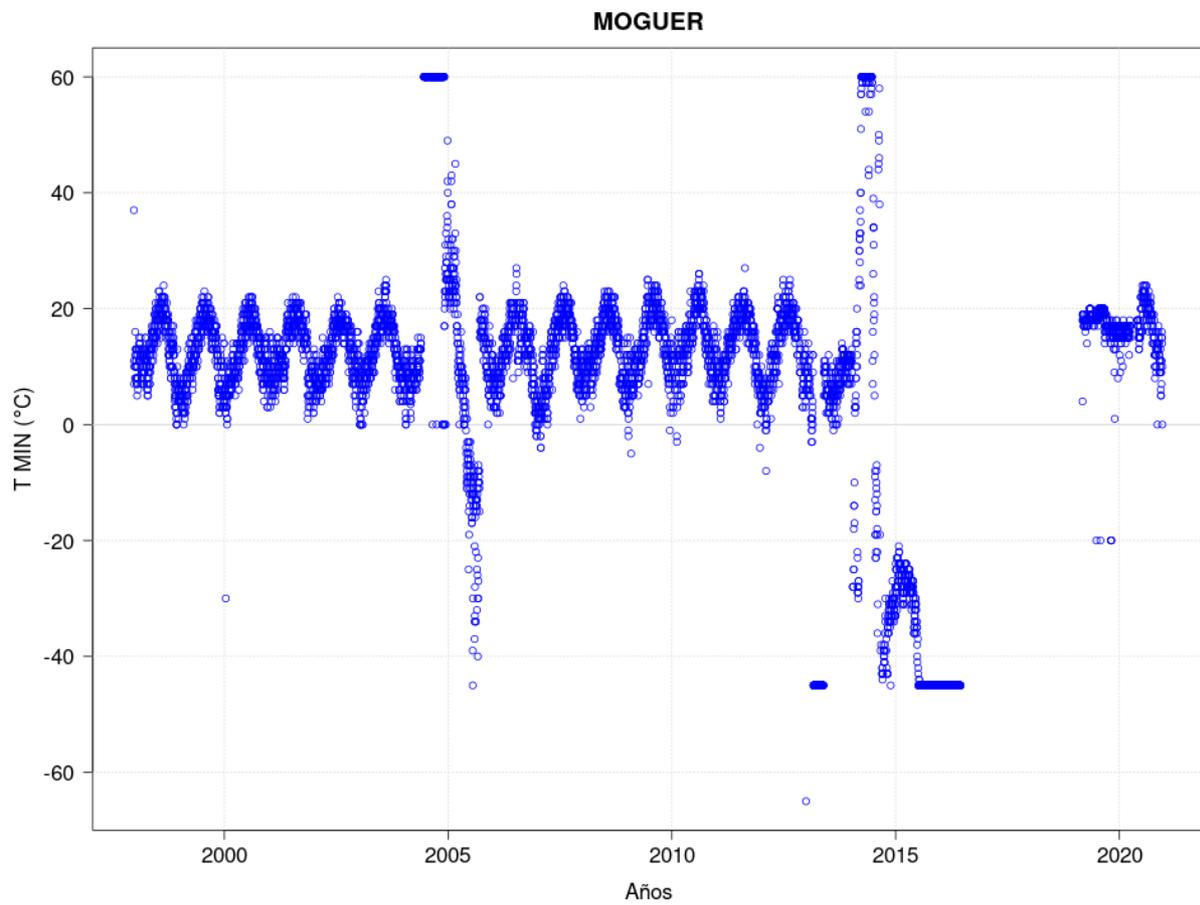


Figura 3.2.5. Representación gráfica de la serie de temperatura mínima diaria del observatorio situado en Moguer. Ejemplo de observatorio eliminado.

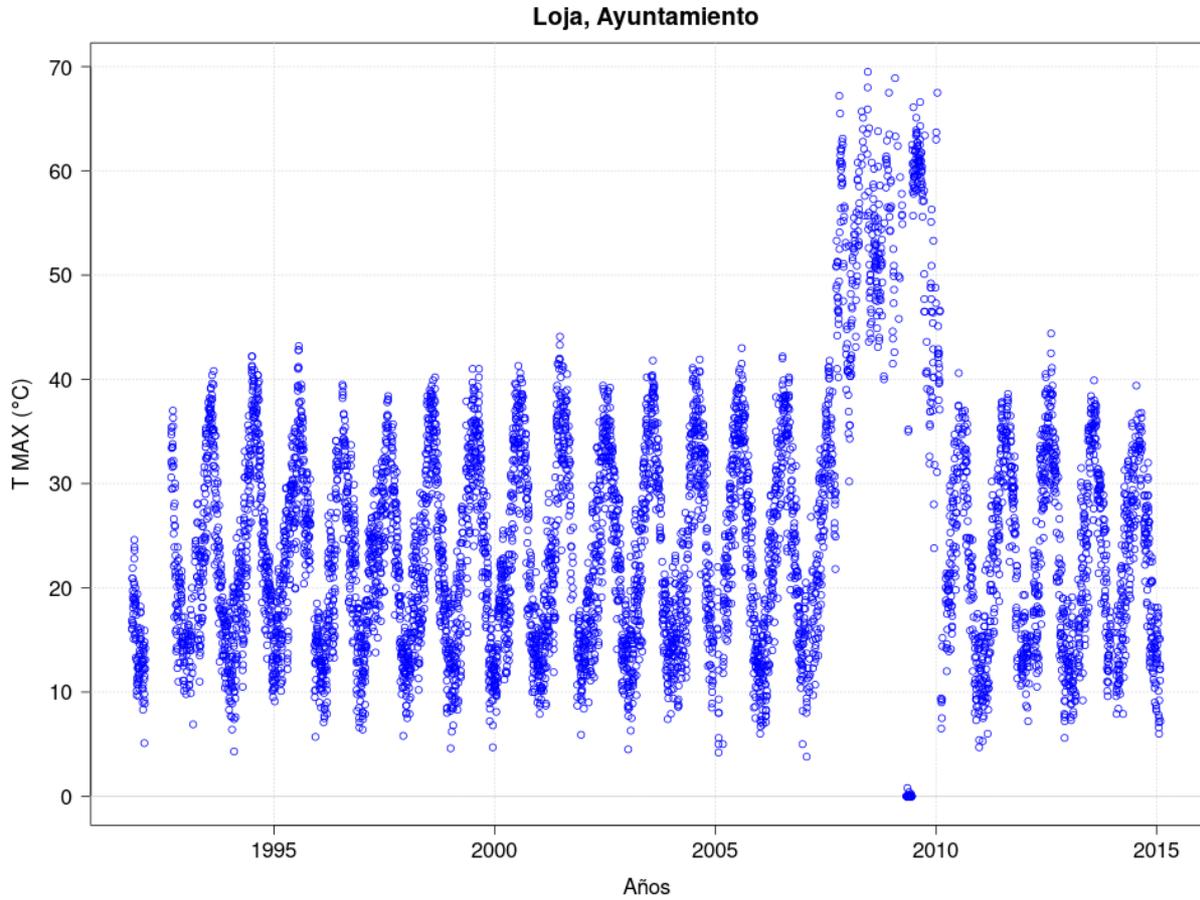


Figura 3.2.6. Representación gráfica de la serie de temperatura máxima diaria del observatorio situado en Loja, Ayuntamiento. Ejemplo de observatorio modificado manualmente eliminando el periodo 2008-2009.

Tabla 3.2.1 Resumen del número de observatorios de temperatura tratados.

Observatorios de temperatura iniciales	1154
Observatorios eliminados por falta de datos antes de proceso de homogeneización	327
Observatorios eliminados al no superar el proceso de homogeneización	8
Observatorios eliminados por duplicado	30
Observatorios de temperatura que han superado todos los controles	789

4. PRECIPITACIÓN

4.1 CONTROL DE CALIDAD

4.1.1. COHERENCIA BÁSICA

Aproximadamente un tercio de los observatorios a analizar (825) presentaban al menos un valor de precipitación negativa en su serie de datos. Se ha realizado un proceso de detección de estos valores y se han sustituido por NA (sin dato).

4.1.2. ELIMINACIÓN DE OUTLIERS

Según el Portal Ambiental de Andalucía, integrado en la Consejería de Agricultura, Ganadería, Pesca y Desarrollo Sostenible de la Junta de Andalucía (https://www.juntadeandalucia.es/medioambiente/portal/web/guest/landing-page-%C3%ADndice/-/asset_publisher/zX2ouZa4r1Rf/content/precipitaci-c3-b3n/20151?categoryVal=), la precipitación máxima diaria para un periodo de retorno de 100 años es de 450mm. Dado que este proyecto se circunscribe al ámbito territorial de Andalucía, se ha escogido un valor de 500mm diarios como límite a partir del cual un valor de la serie histórica es considerado como “outlier”. De esta manera, se ha automatizado un proceso para detectar estos valores y eliminarlos de la serie histórica. Del total de observatorios analizados, 45 presentaban al menos un valor por encima de 500mm. Veamos, a continuación, unos ejemplos.

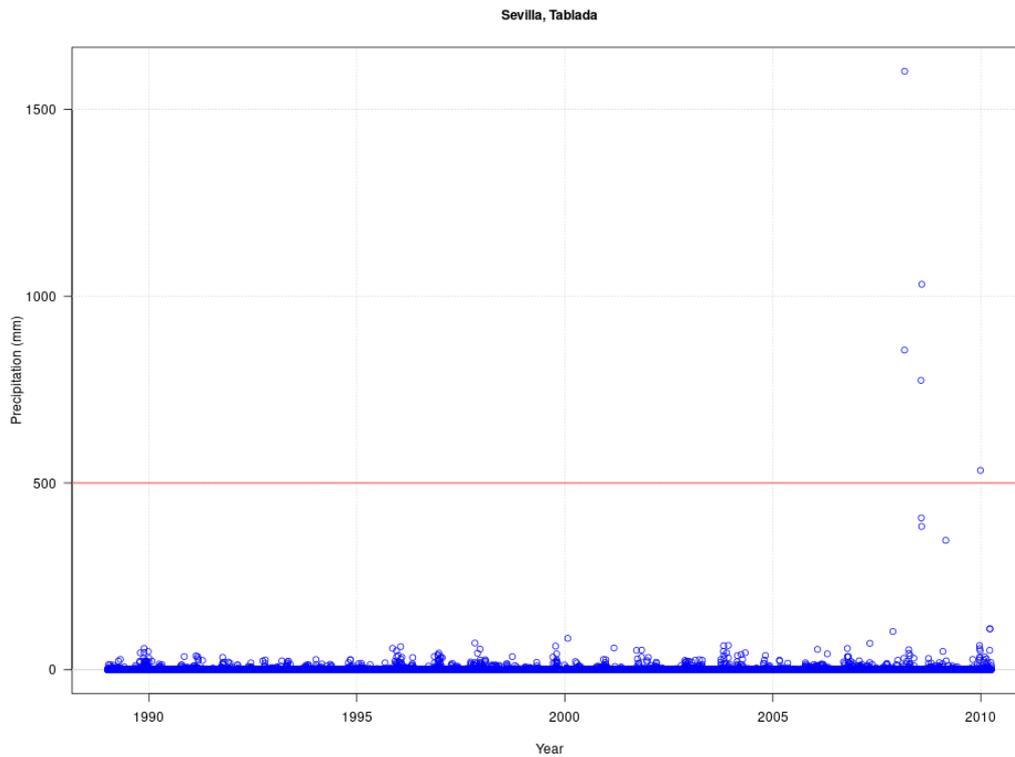


Figura 4.1.2.1. Representación gráfica de la serie de precipitación del observatorio situado en Sevilla-Tablada. Ejemplo de observatorio cuyos outliers han sido detectados y eliminados.

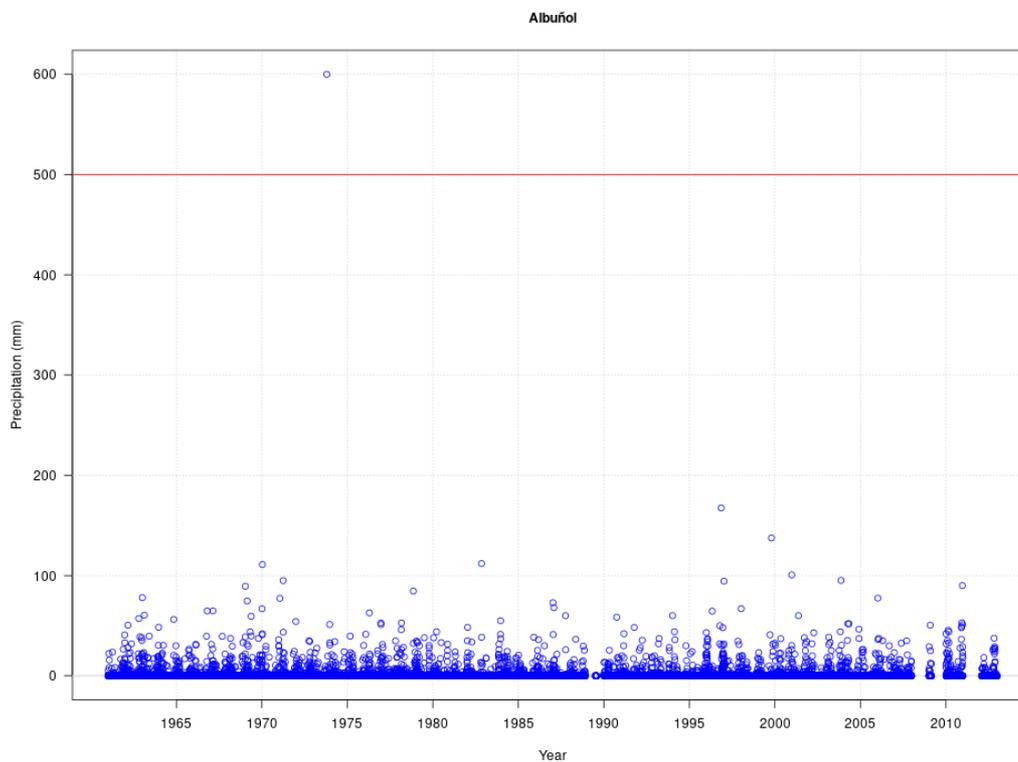


Figura 4.1.2.2. Representación gráfica de la serie de precipitación del observatorio situado en Albuñol. Ejemplo de observatorio cuyos outliers han sido detectados y eliminados.

4.1.3. ELIMINACIÓN DE CEROS

Andalucía tiene un tipo de clima en el que no son extraños largos periodos sin precipitación que pueden superar los 150 días, sobre todo entre mayo y septiembre. El último episodio se produjo en Sevilla en el año 2019, cuando no se registró precipitación en 141 días, entre el 25 de abril y el 13 de septiembre. (https://www.diariodesevilla.es/tiempo-en-Sevilla/tiempo-sevilla-dias-sin-llover_0_1384061804.html)

Dado que no hay estadísticas oficiales sobre número consecutivo de días sin precipitación, se han considerado como buenos los datos de hasta 180 días consecutivos sin precipitación entre los meses de abril y octubre. De esta manera, se han eliminado de las series observadas los conjuntos de más de 180 días consecutivos sin precipitación siempre y cuando ese periodo empiece o acabe entre noviembre y marzo.

De los 1767 observatorios considerados para el estudio, 247 presentan periodos muy largos de precipitación cero, los cuales han sido detectados y eliminados de la serie.

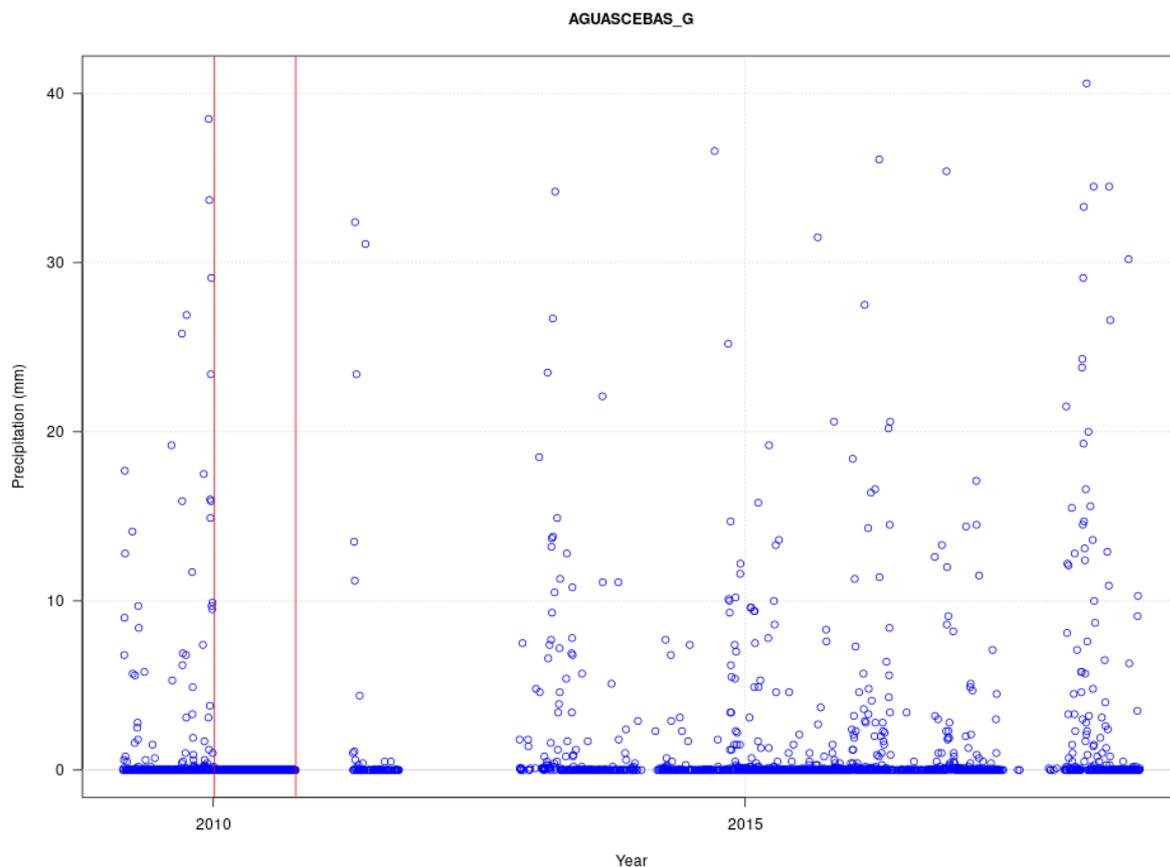


Figura 4.1.3.1. Representación gráfica de la serie de precipitación del observatorio situado en Aguascebas. Ejemplo de observatorio cuyos periodos anómalos de precipitación cero han sido detectados y eliminados.

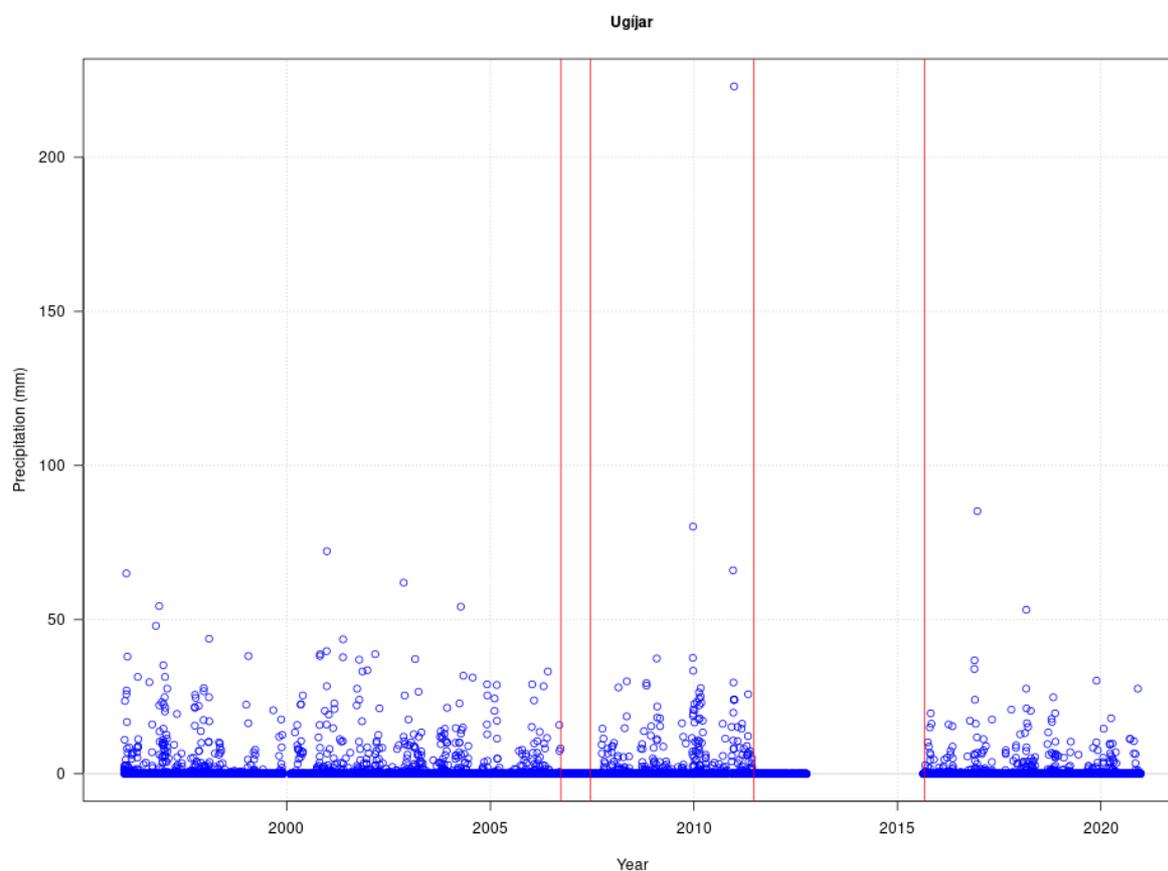


Figura 4.1.3.2. Representación gráfica de la serie de precipitación del observatorio situado en Ugíjar. Ejemplo de observatorio cuyos periodos anómalos de precipitación cero han sido detectados y eliminados.

4.2. HOMOGENEIZACIÓN

Tras haber realizado el control de calidad a las series de precipitación y haber eliminado los outliers y los valores de precipitación cero considerados anómalos, los valores restantes son analizados desde el punto de vista de la homogeneidad. Una vez detectados los saltos de homogeneidad tras aplicar el test de homogeneidad de Kolmogorov-Smirnov, se consideran las siguientes situaciones y acciones:

- Si hay salto de homogeneidad anterior a 1979, la serie no es modificada.
- Si hay al menos un salto de homogeneidad posterior a 1979, se realizará una inspección visual de la serie y se llevarán a cabo las siguientes acciones:
 - Si se comprueba que el periodo de inhomogeneidad se corresponde con un periodo real de precipitación anómala (un año de sequía muy intensa o un año de precipitaciones muy abundantes), no se tendrá en cuenta.

- Si se detecta una única inhomogeneidad, se valorará la eliminación de los datos de la serie que sean anteriores o posteriores a dicha inhomogeneidad si hay un cambio notable en la distribución de la precipitación o si hay un cambio en la precisión del pluviómetro que haga cambiar la distribución.
- Si se detecta más de una inhomogeneidad, se valorará la eliminación de los datos de la serie anteriores o posteriores a cada una de ellas, o la eliminación de los datos comprendidos entre dos inhomogeneidades.
- Si se detectan múltiples inhomogeneidades que llevan aparejados cambios bruscos de más de un orden de magnitud en los valores de precipitación y se intercalan entre periodos sin datos, se eliminará la serie completa.

De los 1767 observatorios que han superado los cribados anteriores, 27 no han superado el proceso de homogeneización y sus series se han descartado por completo. Por otro lado, otros 22 observatorios han sufrido algún cambio en sus series debido a la eliminación de algún conjunto de datos por inhomogeneidades. De esta manera, hay 1740 observatorios de precipitación que han superado los controles y cuyas series podrán ser utilizadas para el estudio.

A continuación se muestran unos ejemplos de cómo se ha procedido en el proceso de homogeneización

En este primer caso, al igual que en muchos de los observatorios analizados, se ha detectado una inhomogeneidad en el año 2010. Pero en la mayoría de los casos, esta inhomogeneidad tiene una explicación: finales de 2009 y principios de 2010 fue un periodo excepcionalmente lluvioso, sobre todo en la cuenca del Guadalquivir, por ello se ha decidido no alterar las series que presenten una inhomogeneidad en este periodo.

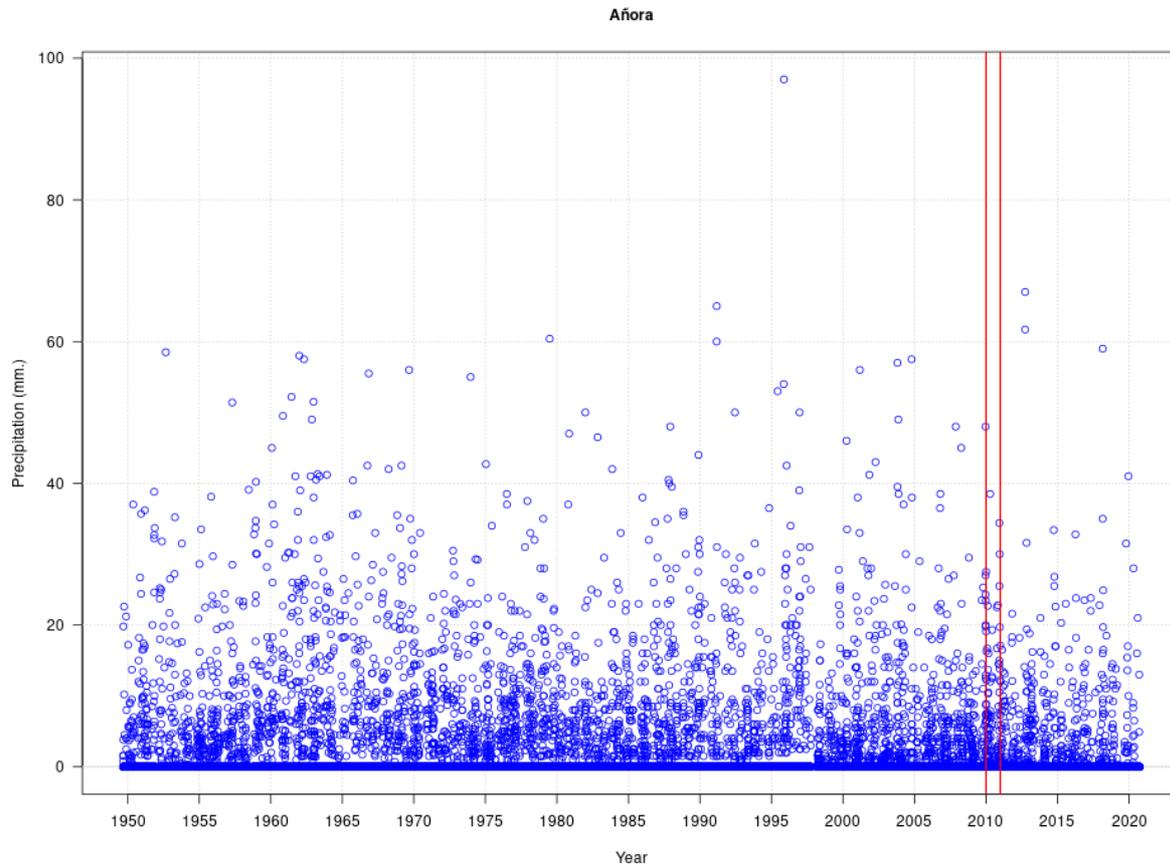


Figura 4.2.1. Representación gráfica de la serie de precipitación del observatorio situado en Añora. Ejemplo de observatorio que presenta una inhomogeneidad debido a un periodo excepcionalmente lluvioso.

En el siguiente caso, distinguen tres periodos distintos: un periodo anterior a 2004, donde los datos son erróneos a simple vista ya que sólo hay tres posibles valores de precipitación (0, 200 y 400 mm), otro periodo entre 2007 y 2017 donde, a pesar de que la precipitación aparentemente tiene una distribución correcta, sus valores son un orden de magnitud superiores a lo esperable, y un tercer periodo a partir de 2017 donde parece que los datos son correctos. Ante este escenario, y teniendo en cuenta que el número de datos correctos es inferior a 2000, se ha decidido eliminar la serie para el estudio.

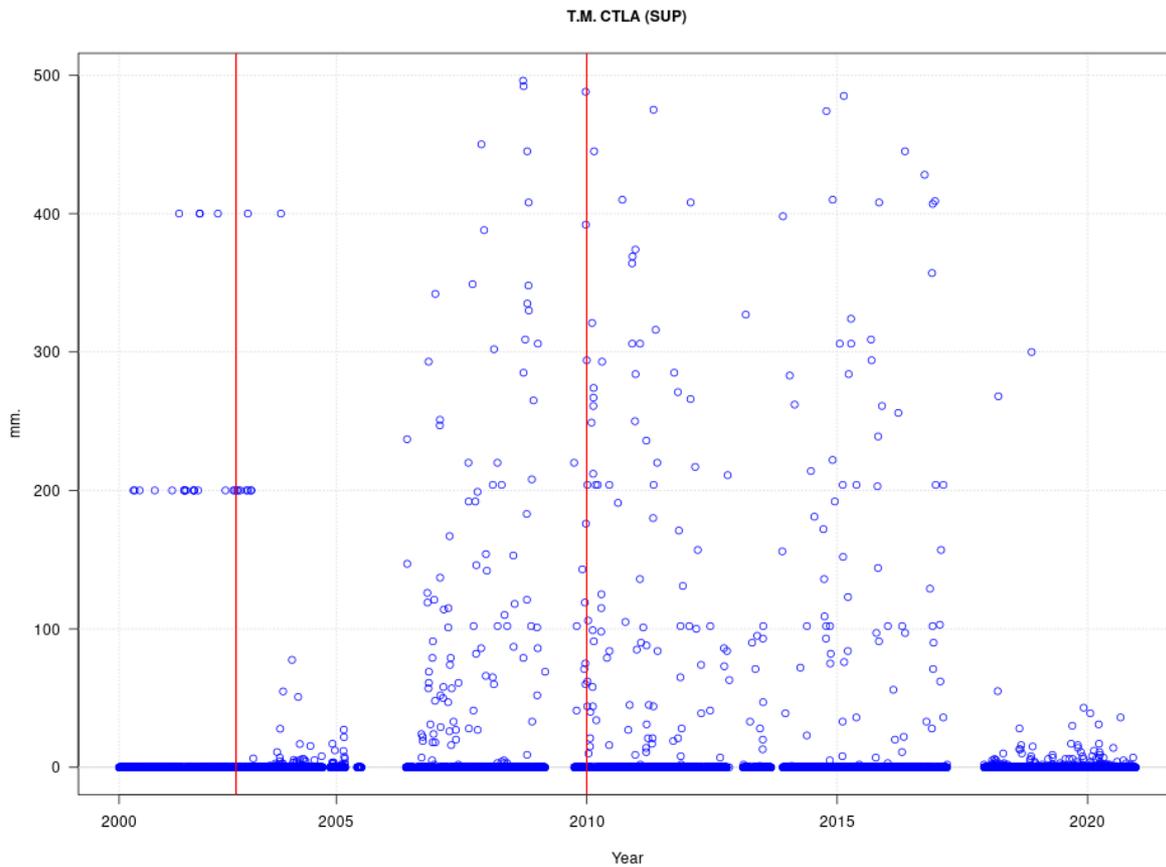


Figura 4.2.2. Representación gráfica de la serie de precipitación del observatorio situado en Carboneras. Ejemplo de observatorio cuya serie se ha descartado por completo debido a las múltiples inhomogeneidades que presenta.

En el siguiente ejemplo, se observa que hay dos periodos bien diferenciados, separados por una inhomogeneidad detectada en el año 2010. En este caso, parece que los datos anteriores a 2010 son un orden de magnitud superior a lo esperado, posiblemente debido a una lectura errónea de la precisión del pluviómetro que hace cambiar totalmente la distribución de la precipitación a uno y otro lado de ese año. De esta manera, se ha procedido a eliminar los datos de la serie anteriores a 2010:

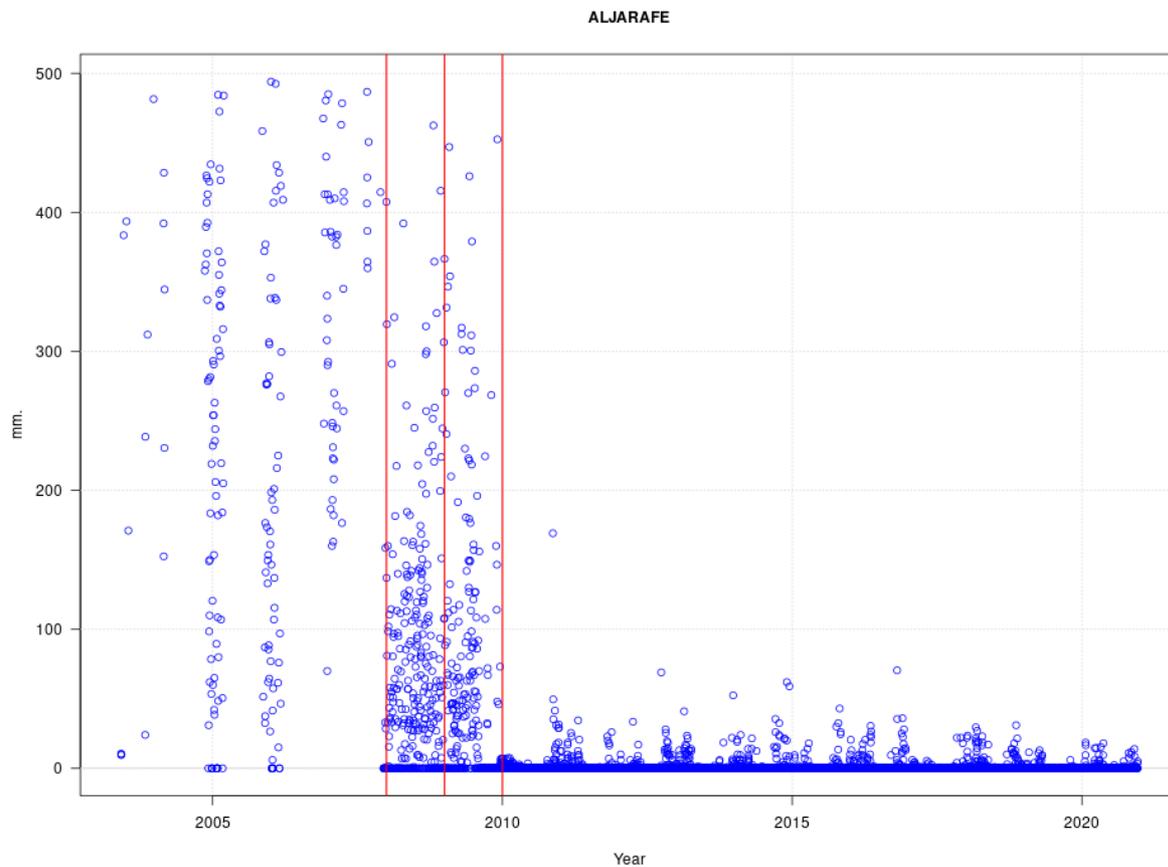


Figura 4.2.3. Representación gráfica de la serie de precipitación del observatorio situado en Aljarafe. Ejemplo de observatorio en el que se ha eliminado un tramo de la serie.

En el siguiente caso, el programa ha detectado una inhomogeneidad en el año 2014. Al contrario que en los casos anteriores, el error en la serie parece estar en el último tramo de la misma, y por presentar valores más bajos que en el resto de la serie histórica. Lo que se ha decidido en este caso ha sido eliminar de la serie todos los datos posteriores a 2014:

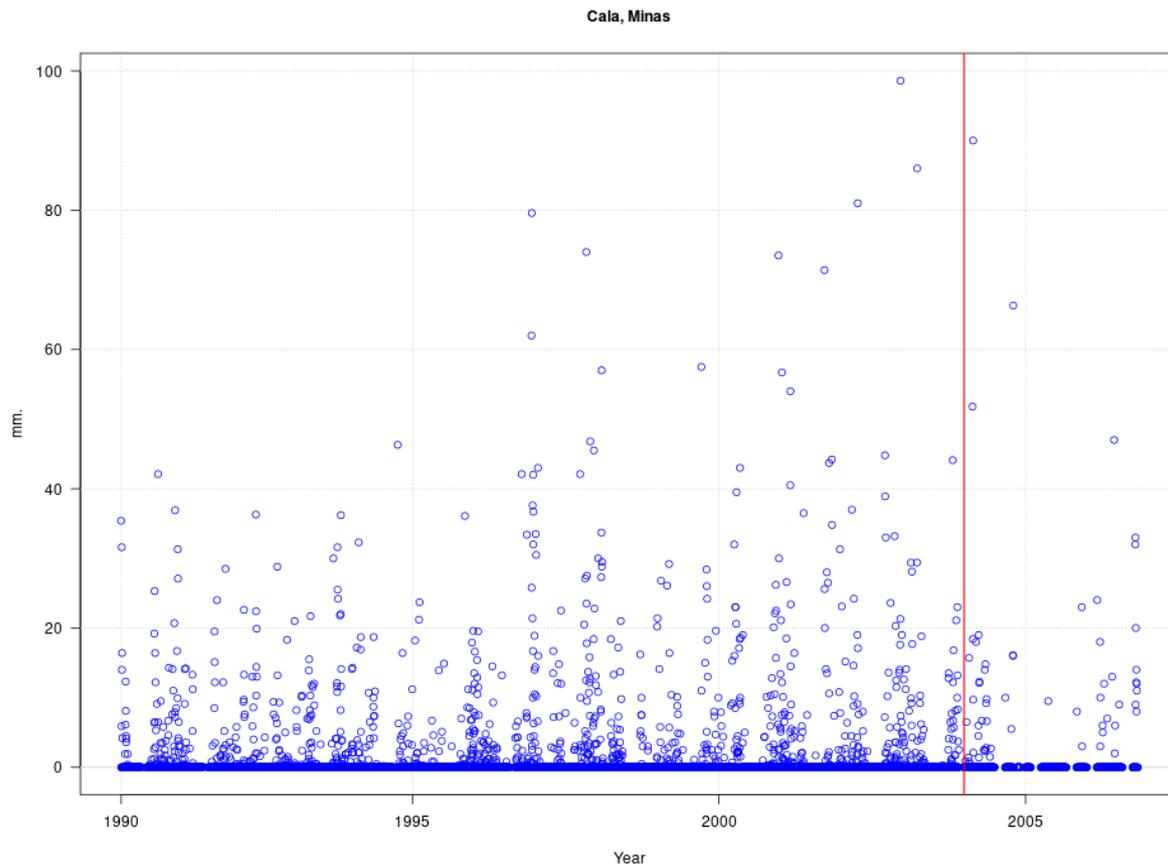


Figura 4.2.4. Representación gráfica de la serie de precipitación del observatorio situado en Cala. Ejemplo de observatorio en el que se ha eliminado una parte de la serie.

Una vez concluido los procesos de control de calidad y de homogeneización, se ha vuelto a ejecutar el proceso que cuenta el número de datos diarios de la serie entre los años 1979 y 2020, y ha detectado que otros 16 observatorios han caído por debajo de los 2000 datos diarios al eliminar una parte de sus series. De esta forma, de los 2463 observatorios iniciales con serie de precipitación, se utilizarán en el estudio 1724.

Tabla 4.2.1 Resumen del número de observatorios de precipitación tratados.

Observatorios de precipitación iniciales	2463
Observatorios eliminados por falta de datos antes del proceso de homogeneización	696
Observatorios eliminados al no superar el proceso de homogeneización	27
Observatorios eliminados por falta de datos después del proceso de homogeneización	16
Observatorios de precipitación que han superado todos los controles	1724