JORGE LUIS GONZALEZ CASTRO

CC: 1032395475

https://github.com/jlgingcivil/R.LTWB.CS2120

r.ltwb – SECTION 03

Descarga, procesamiento y análisis de datos hidroclimatológicos

dentificación y procesamiento de datos atípicos - Outliers

**TABLA DE CONTENIDO**

[1. Introducción 3](#_Toc145688373)

[2. Objetivo General 3](#_Toc145688374)

[3. Actividad 1: Procesamiento en software 3](#_Toc145688375)

[4. ACTIVIDAD 2: análisis de OTROS PARÁMETROS 3](#_Toc145688376)

[5. ACTIVIDAD 3: SENSIBILIDAD PARÁMETROS ESTADÍSTICOS 3](#_Toc145688377)

[6. Conclusiones 11](#_Toc145688378)

[7. Referencias Bibliográficas 1](#_Toc145688379)

**ÍNDICE DE ILUSTRACIONES**

[Ilustración 3‑1. Descarga Outlier.py 3](#_Toc145688380)

[Ilustración 3‑2. Script Outlier.py 4](#_Toc145688381)

[Ilustración 3‑3. Ejecución inicial Outlier.py 4](#_Toc145688382)

[Ilustración 3‑4. Resultados Precipitación script Outlier.py 6](#_Toc145688383)

[Ilustración 3‑5. Resultados Outlier Método 1 Precipitación 7](#_Toc145688384)

[Ilustración 3‑6. Resultados Outlier Método 2 Precipitación 7](#_Toc145688385)

[Ilustración 3‑7. Resultados Outlier Método 3 Precipitación 8](#_Toc145688386)

[Ilustración 3‑8. Resultados Outlier Método 1 Evaporación 8](#_Toc145688387)

[Ilustración 3‑9. Resultados Outlier Método 2 Evaporación 9](#_Toc145688388)

[Ilustración 3‑10. Resultados Outlier Método 3 Evaporación 9](#_Toc145688389)

[Ilustración 3‑11. Resultados Outlier Método 1 Temperatura Máxima 10](#_Toc145688390)

[Ilustración 3‑12. Resultados Outlier Método 2 Temperatura Máxima 10](#_Toc145688391)

[Ilustración 3‑13. Resultados Outlier Método 3 Temperatura Máxima 11](#_Toc145688392)

[Ilustración 3‑14. Resultados Outlier Método 1 Temperatura Mínima 11](#_Toc145688393)

[Ilustración 3‑15. Resultados Outlier Método 2 Temperatura Mínima 12](#_Toc145688394)

[Ilustración 3‑16. Resultados Outlier Método 3 Temperatura Mínima 12](#_Toc145688395)

[Ilustración 3‑17. Resultados Outlier Método 1 Caudal 13](#_Toc145688396)

[Ilustración 3‑18. Resultados Outlier Método 2 Caudal 13](#_Toc145688397)

[Ilustración 3‑19. Resultados Outlier Método 3 Caudal 14](#_Toc145688398)

[Ilustración 3‑20. Registros Evaporación 14](#_Toc145688399)

[Ilustración 3‑21. Análisis Desviaciones Estándar Precipitación 15](#_Toc145688400)

[Ilustración 3‑22. Registros Evaporación 1](#_Toc145688401)

[Ilustración 3‑23. Análisis Desviaciones Estándar Evaporación 1](#_Toc145688402)

[Ilustración 3‑24. Registros Caudal 1](#_Toc145688403)

[Ilustración 3‑25. Análisis Desviaciones Estándar Caudal 1](#_Toc145688404)

[Ilustración 3‑26. Registros Temperatura Mínima 2](#_Toc145688405)

[Ilustración 3‑27. Registros Temperatura Máxima 2](#_Toc145688406)

[Ilustración 4‑1. Acceso plataforma IDEAM 3](#_Toc145688407)

[Ilustración 5‑1. Resultados Precipitación 4](#_Toc145688408)

[Ilustración 5‑2. Resultados Evaporación 5](#_Toc145688409)

[Ilustración 5‑3. Resultados Caudal 7](#_Toc145688410)

[Ilustración 5‑4. Resultados Temperatura Máxima 8](#_Toc145688411)

[Ilustración 5‑5. Resultados Temperatura Mínima 9](#_Toc145688412)

# Introducción

Se continua con curso Balance hidrológico de largo plazo para estimación de caudales medios usando SIG – LWTB con el desarrollo de la sección 3 Descarga, procesamiento y análisis de datos hidroclimatológicos. A continuación, se presenta en cada numeral las actividades realizadas de acuerdo con cada capítulo de la sección de estudio, incluyendo el resumen de actividades, logros alcanzados y capturas de pantalla de los ejercicios realizados. Se ha creado el repositorio <https://github.com/jlgingcivil/R.LTWB.CS2021> para la inclusión de los archivos y documentos de las actividades desarrolladas.

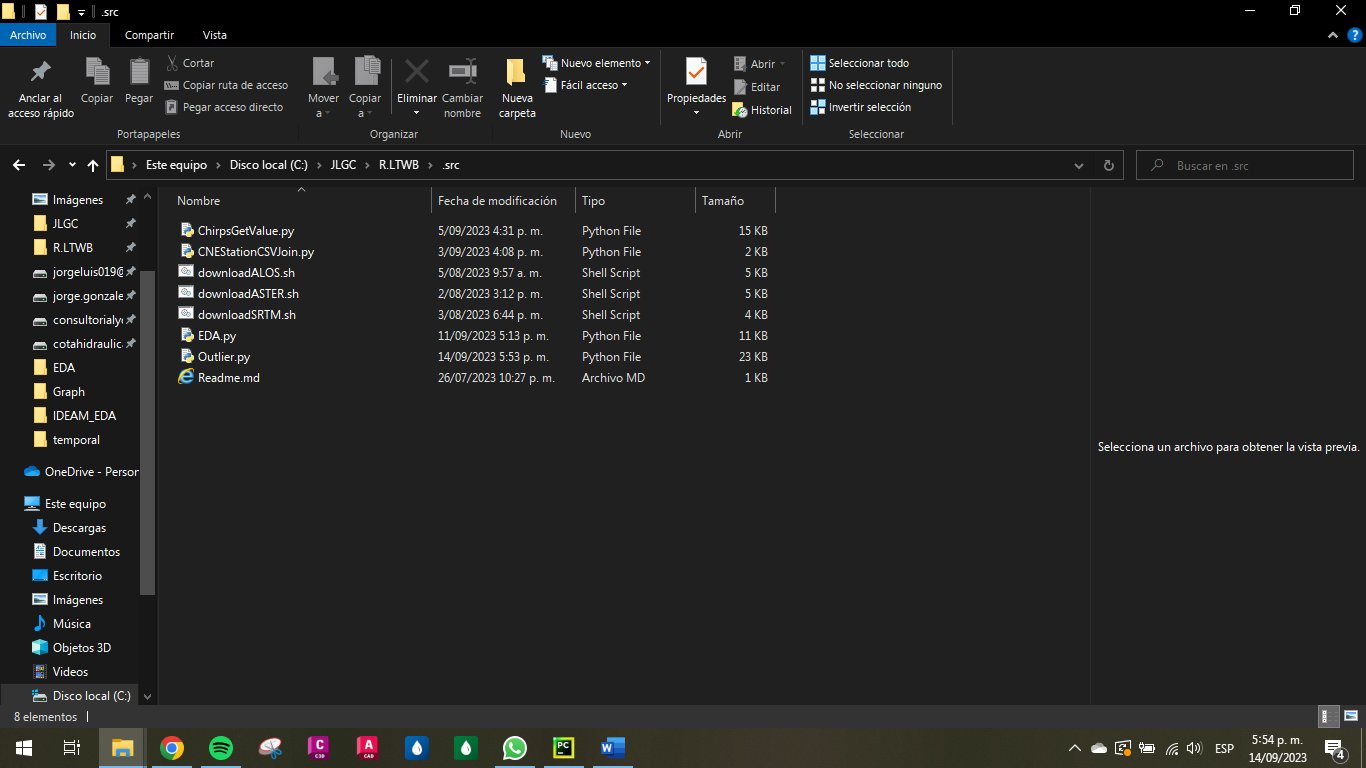
# Objetivo General

El objetivo general en esta sección es analizar a través de diferentes técnicas los datos atípicos, realizar su exclusión, complemento o ajuste de las series de datos descargadas desde el portal del IDEAM.

# Actividad 1: Procesamiento en software

En primera medida se realiza la descarga del script Outlier.py y la creación de la carpeta para almacenamiento de archivos.

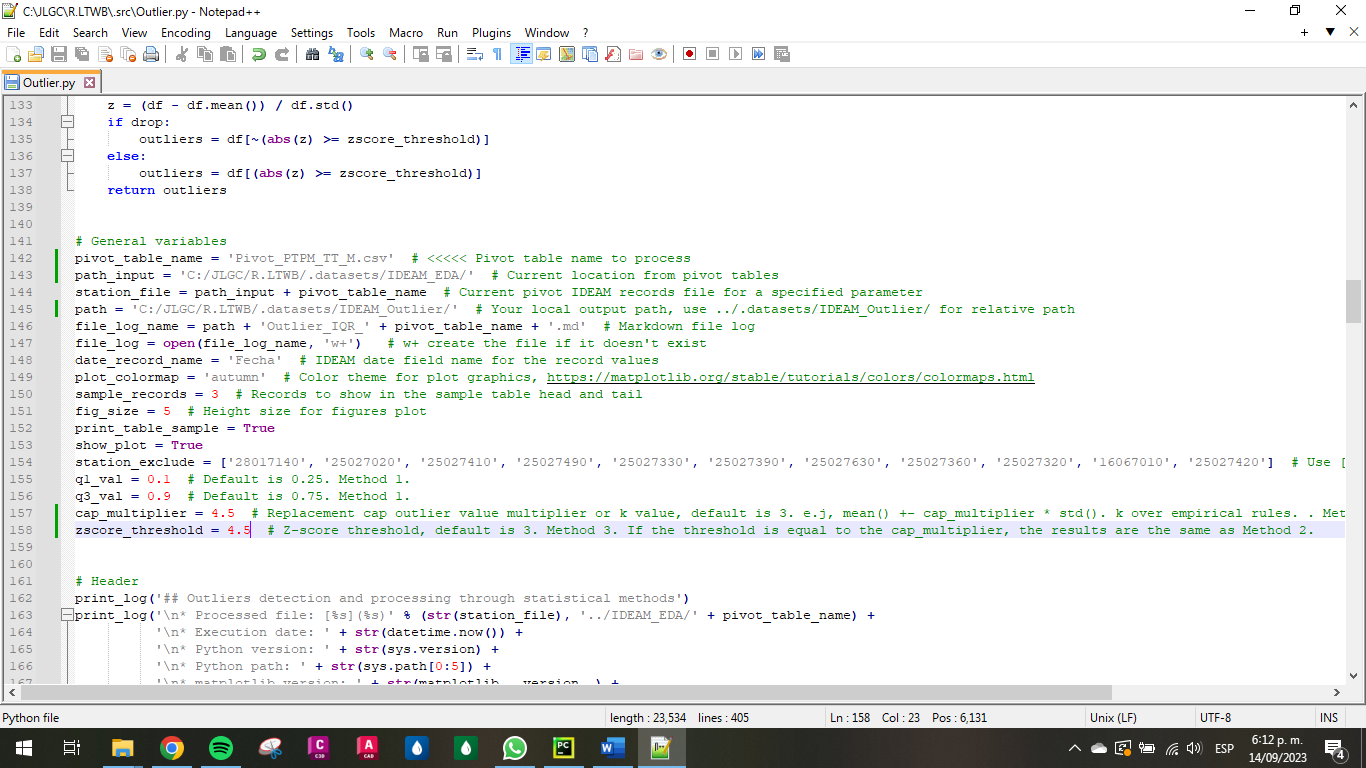
Ilustración ‑. Descarga Outlier.py



Fuente: Elaboración Propia, 2023.

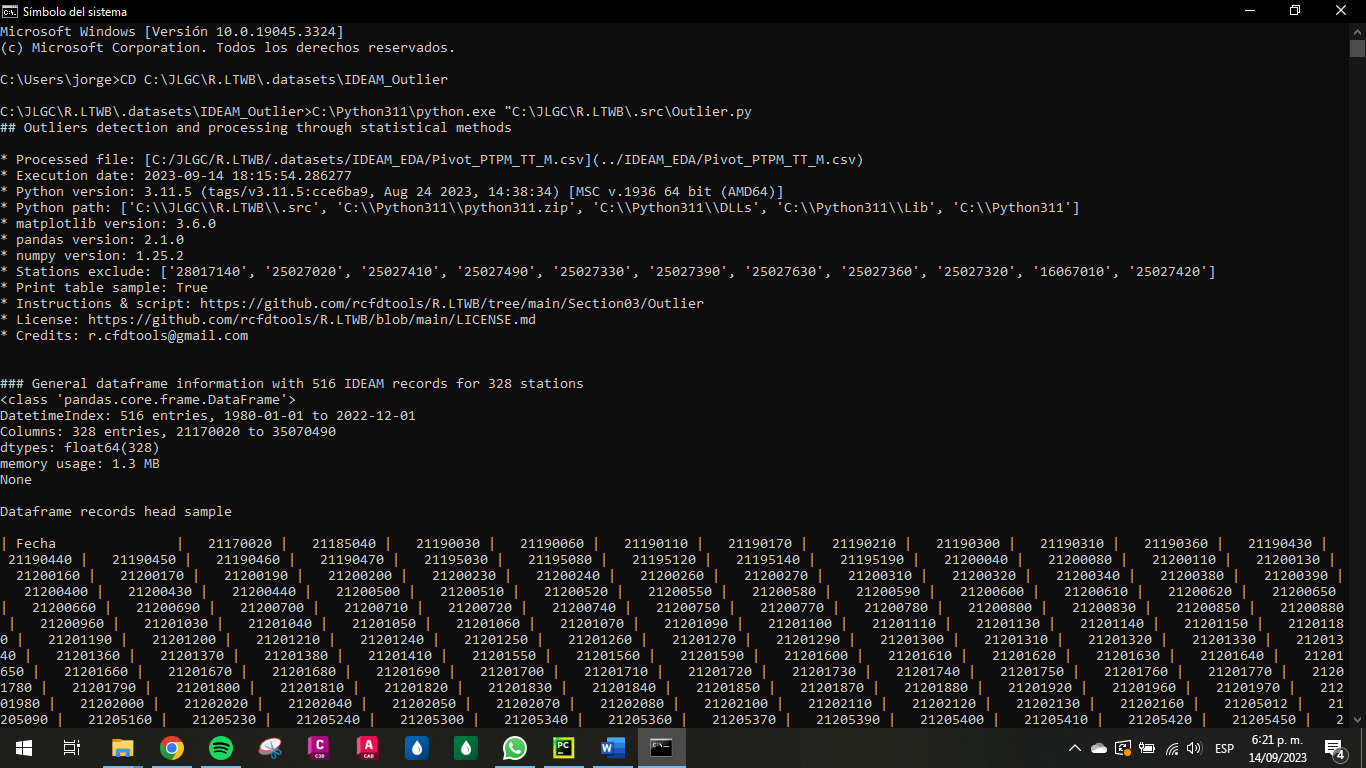
Luego se ajusta el script para que lea la ruta de almacenamiento de archivos, así como la definición de los parámetros tomando el ejemplo de clase para el archivo pivot de precipitación y se inicia con la ejecución de la herramienta.

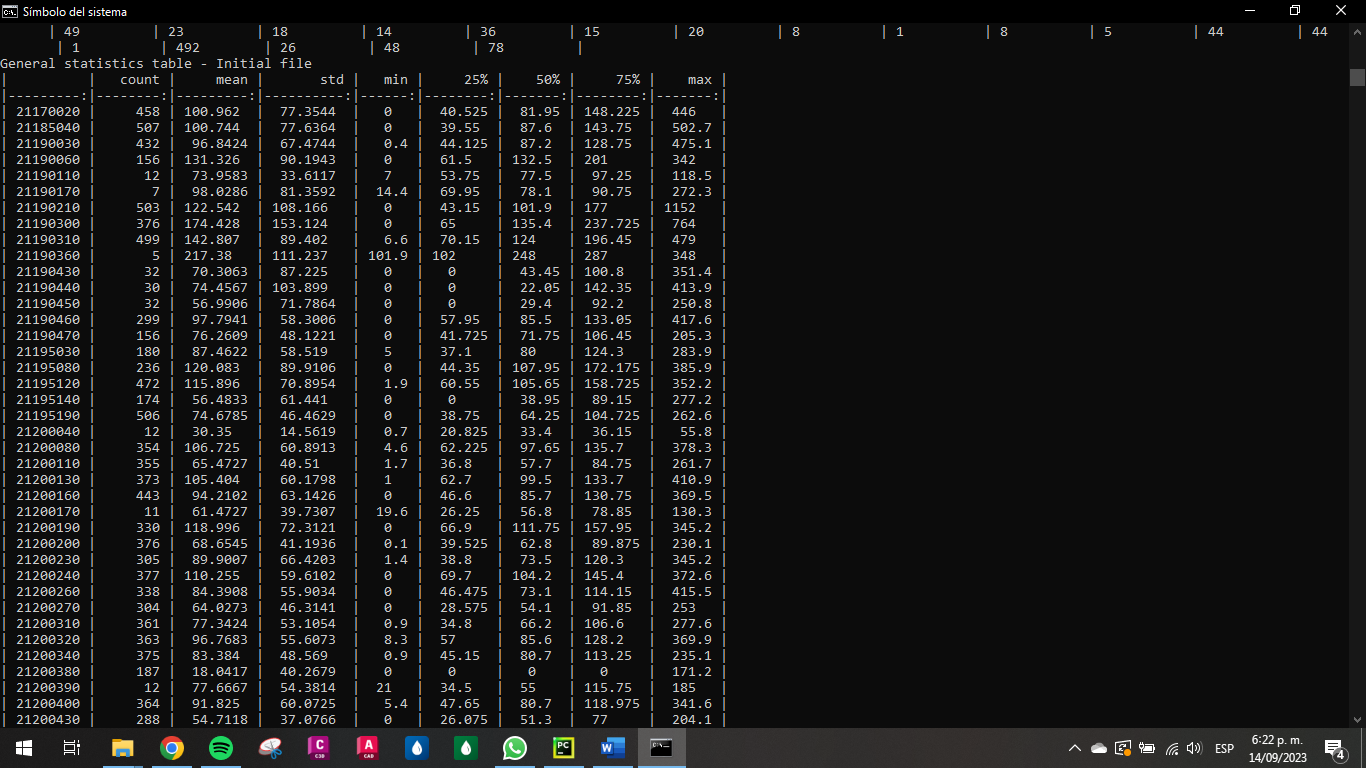
Ilustración ‑. Script Outlier.py

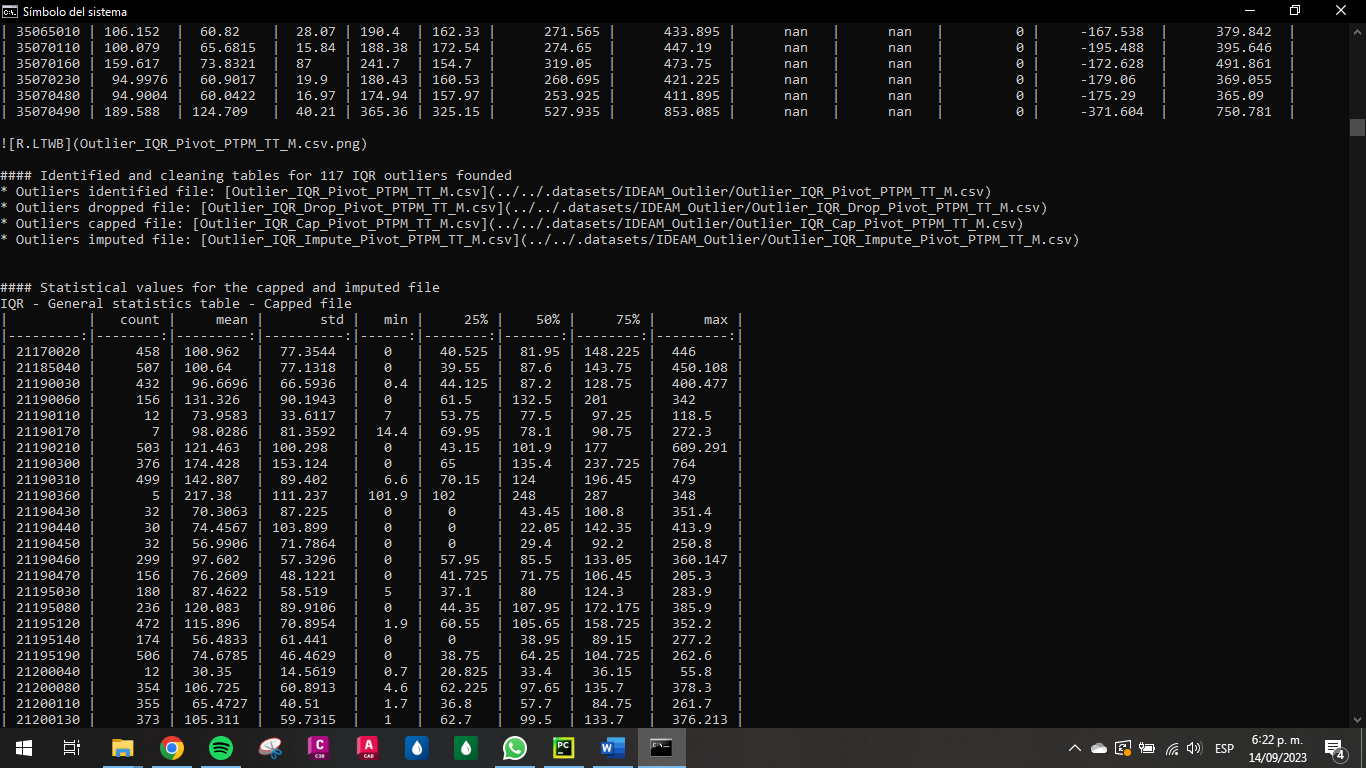


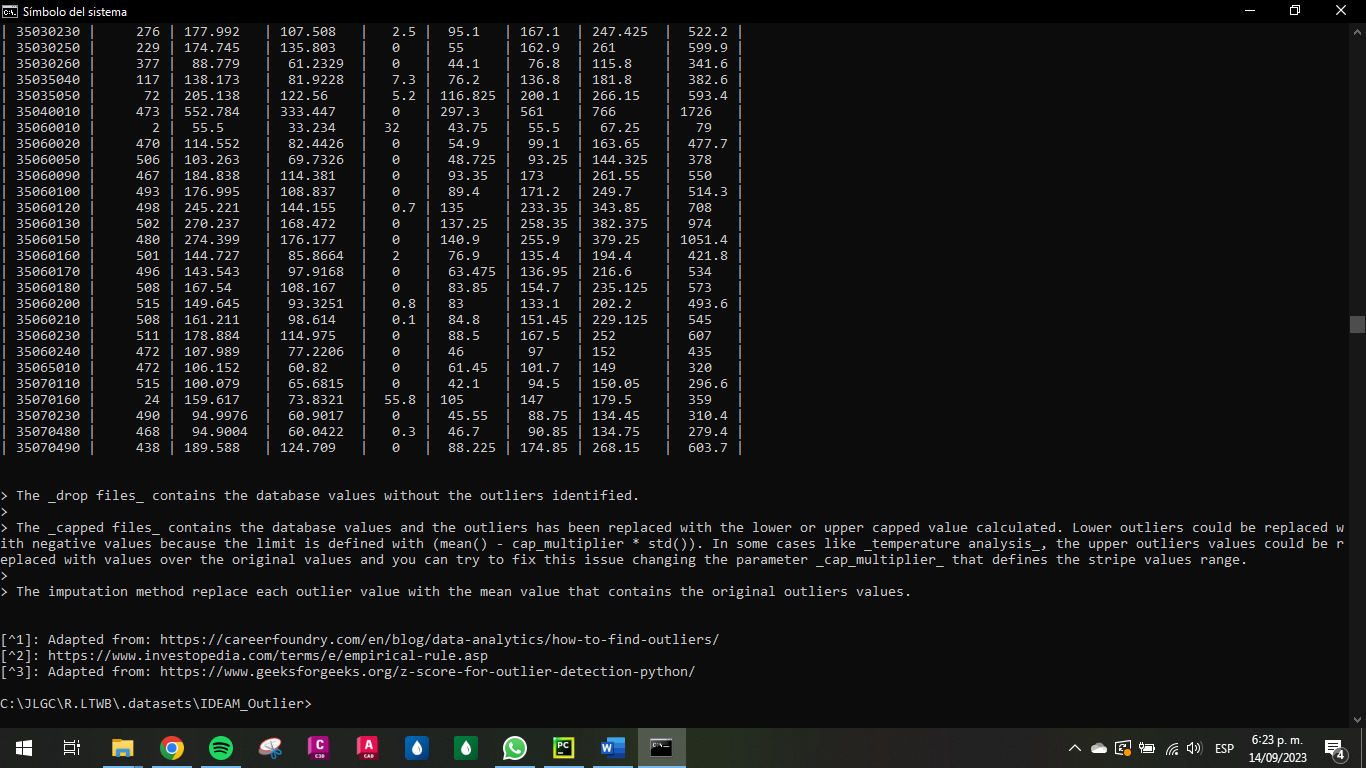
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Ejecución inicial Outlier.py





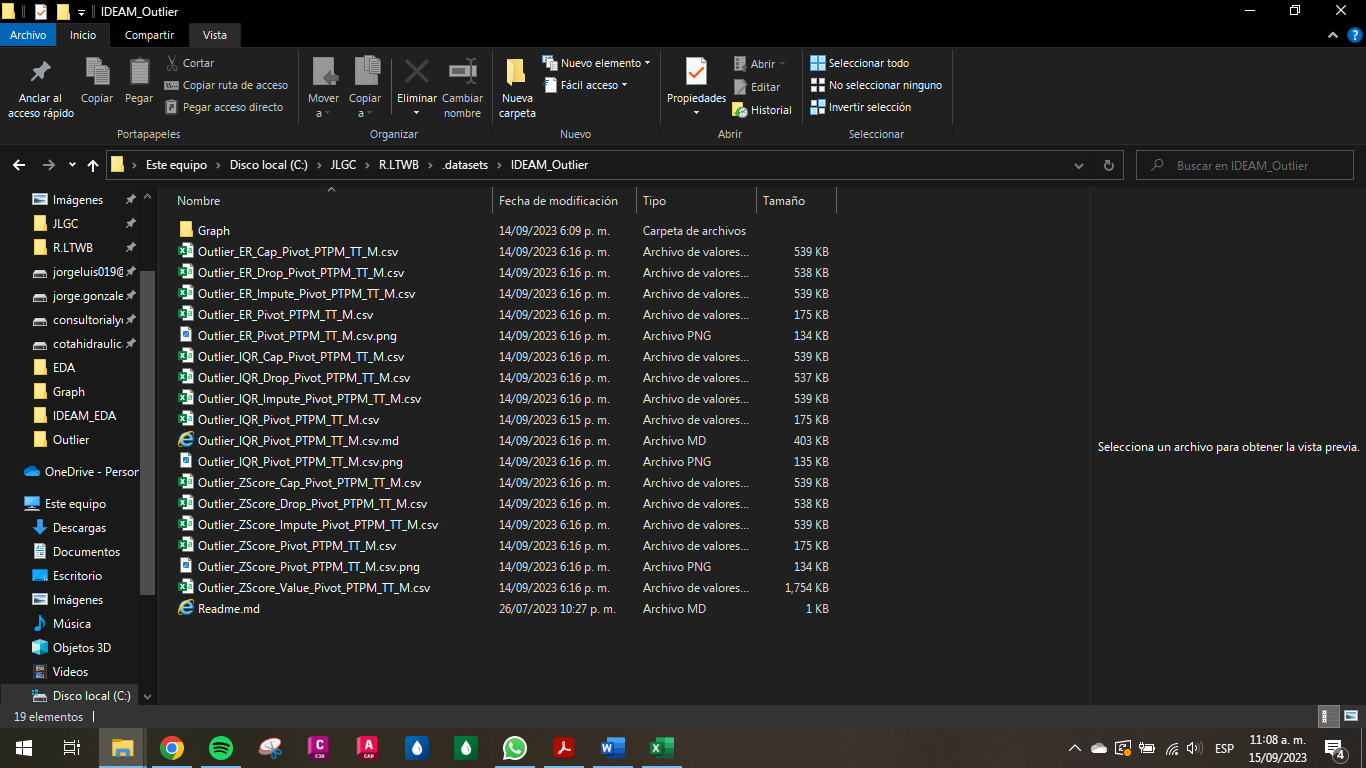




Fuente: Elaboración Propia, 2023.

Se verificó que en la carpeta /.datasets/IDEAM\_Outlier se almacenarán los resultados del script en cuanto a gráficas, tablas y archivo de visualización en formato MarkDown.

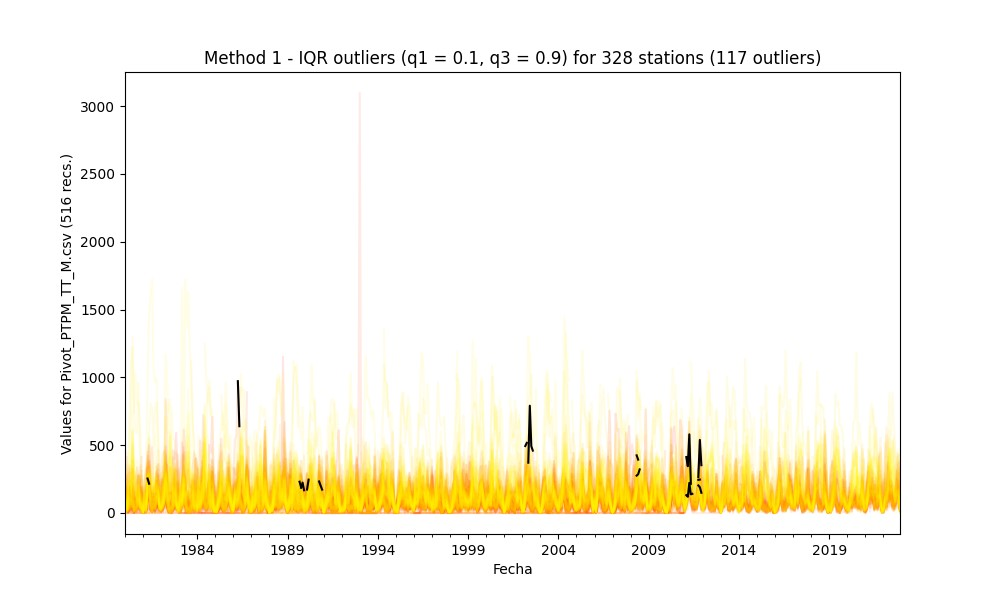
Ilustración ‑. Resultados Precipitación script Outlier.py



Fuente: Elaboración Propia, 2023.

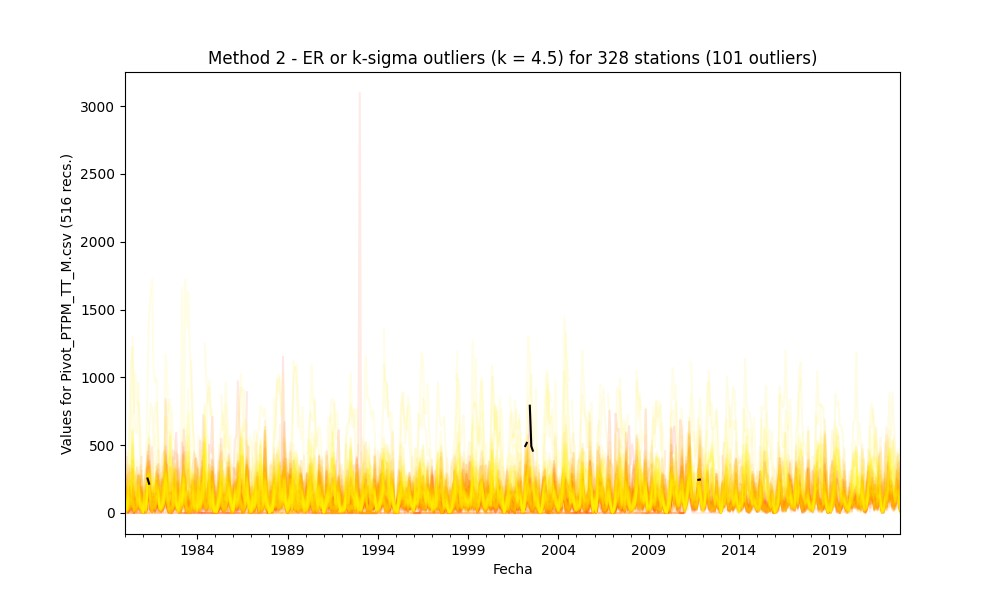
Las gráficas de resultados por cada método se presentan a continuación.

Ilustración ‑. Resultados Outlier Método 1 Precipitación



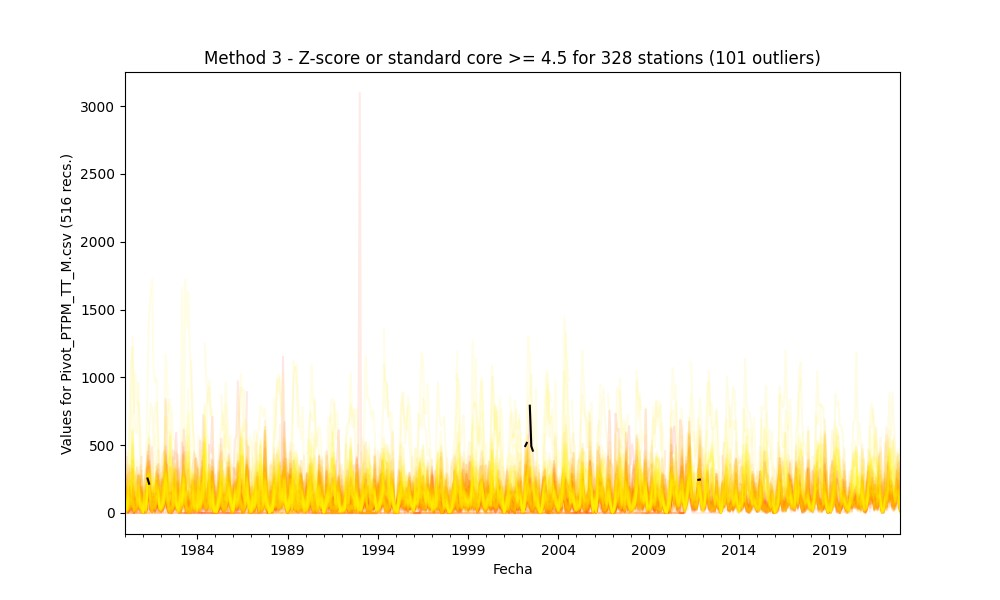
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Resultados Outlier Método 2 Precipitación



Fuente: Elaboración Propia, 2023.

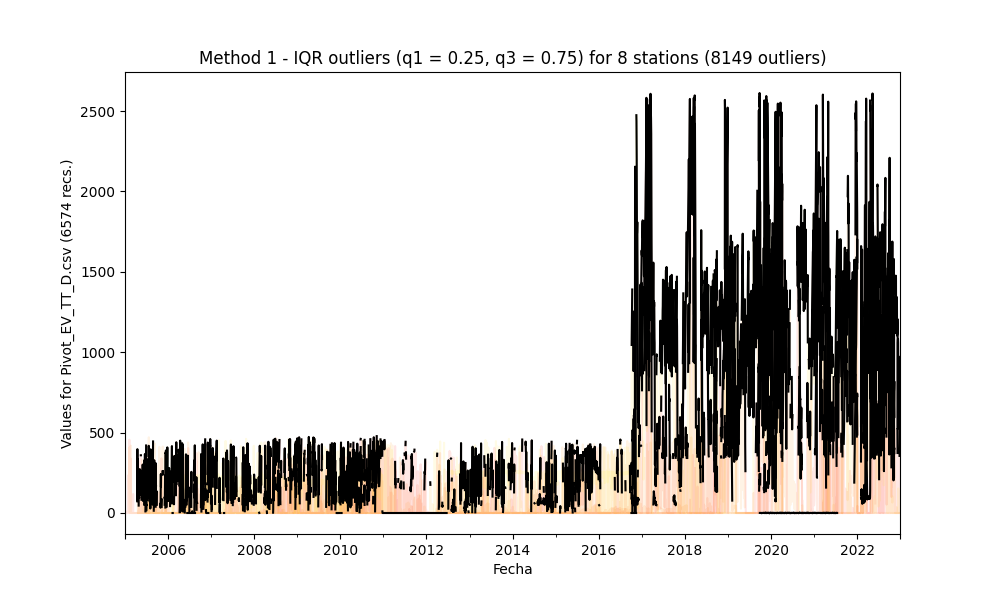
Ilustración ‑. Resultados Outlier Método 3 Precipitación



Fuente: Elaboración Propia, 2023.

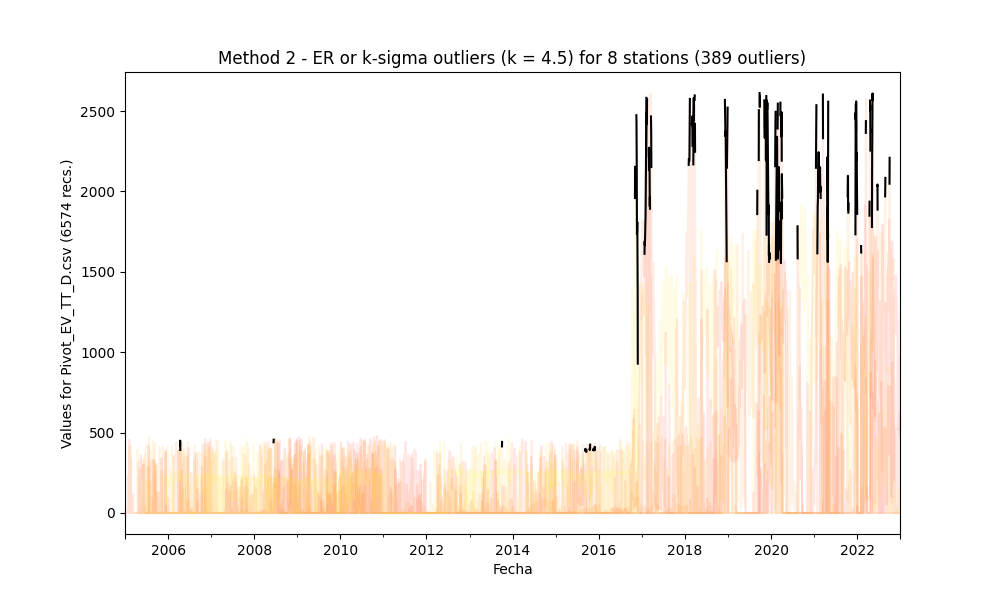
Se realizó el mismo ejercicio para los parámetros de caudal, evaporación, temperatura máxima y temperatura mínima.

Ilustración ‑. Resultados Outlier Método 1 Evaporación



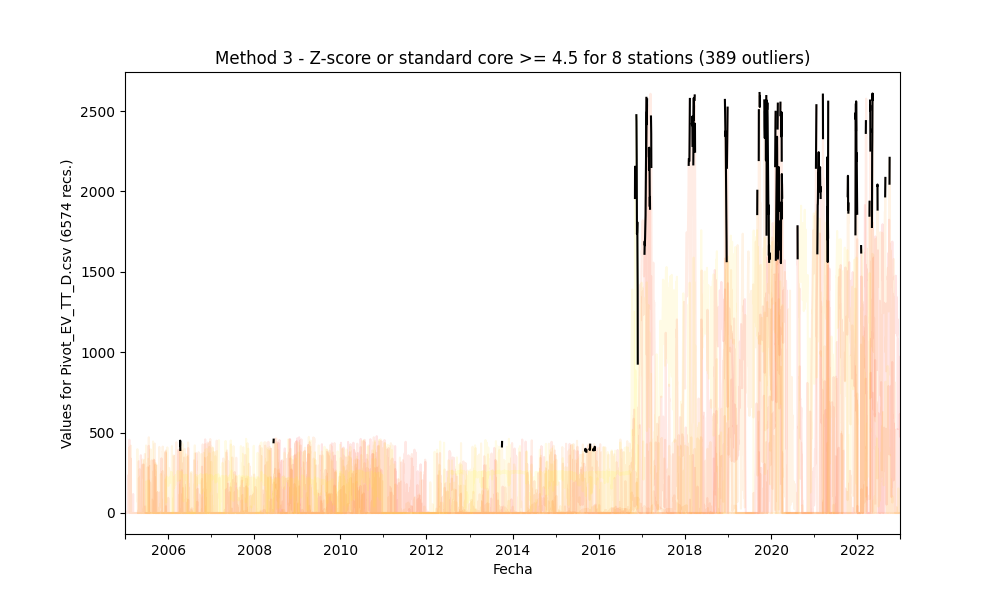
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Resultados Outlier Método 2 Evaporación



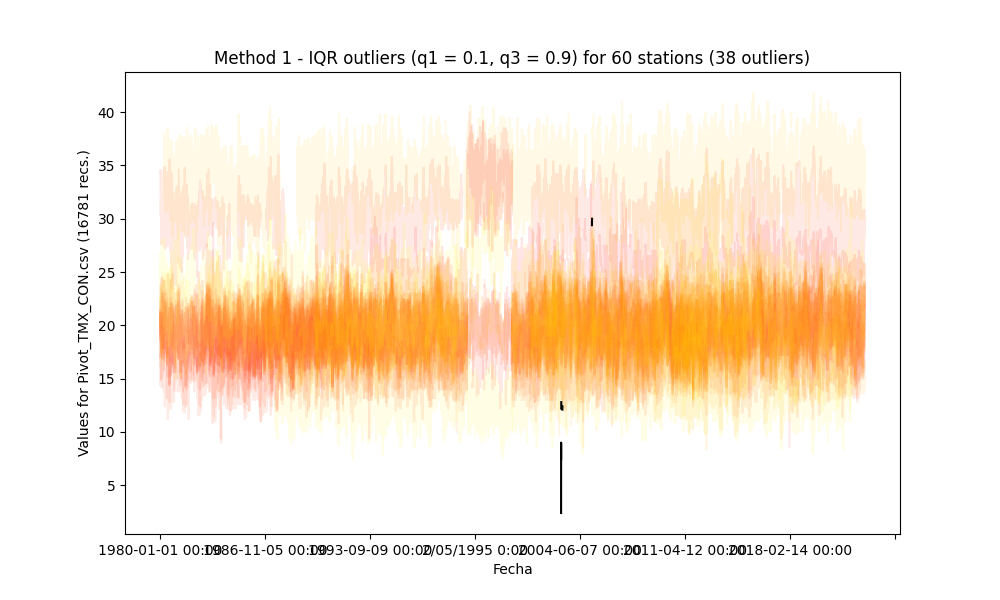
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Resultados Outlier Método 3 Evaporación



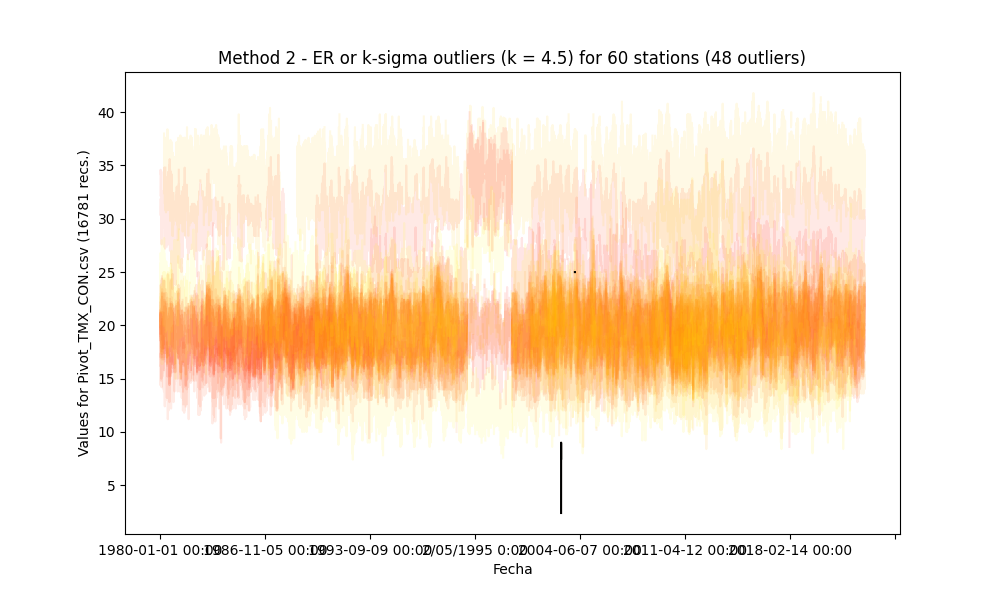
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Resultados Outlier Método 1 Temperatura Máxima



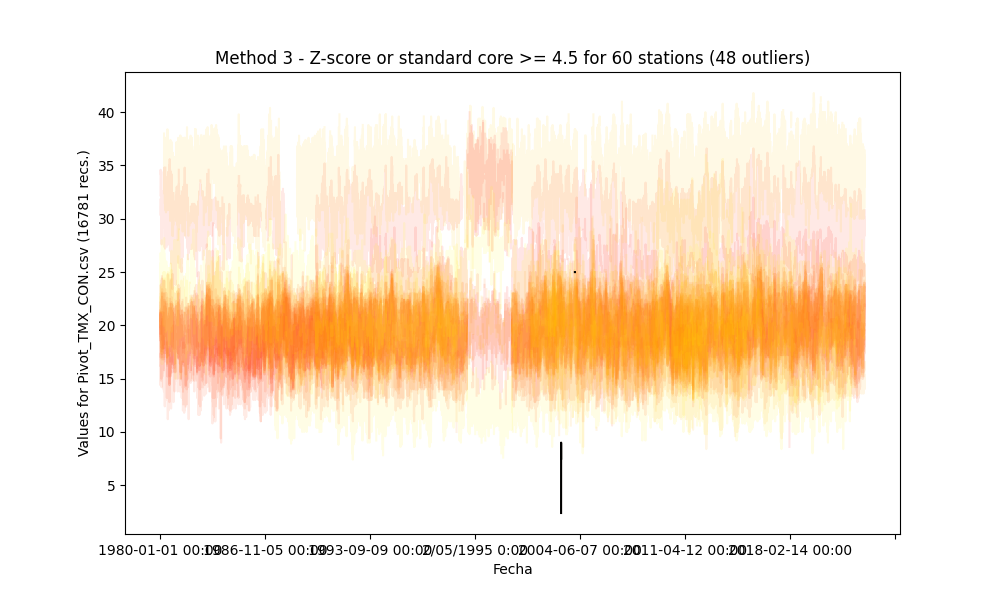
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Resultados Outlier Método 2 Temperatura Máxima



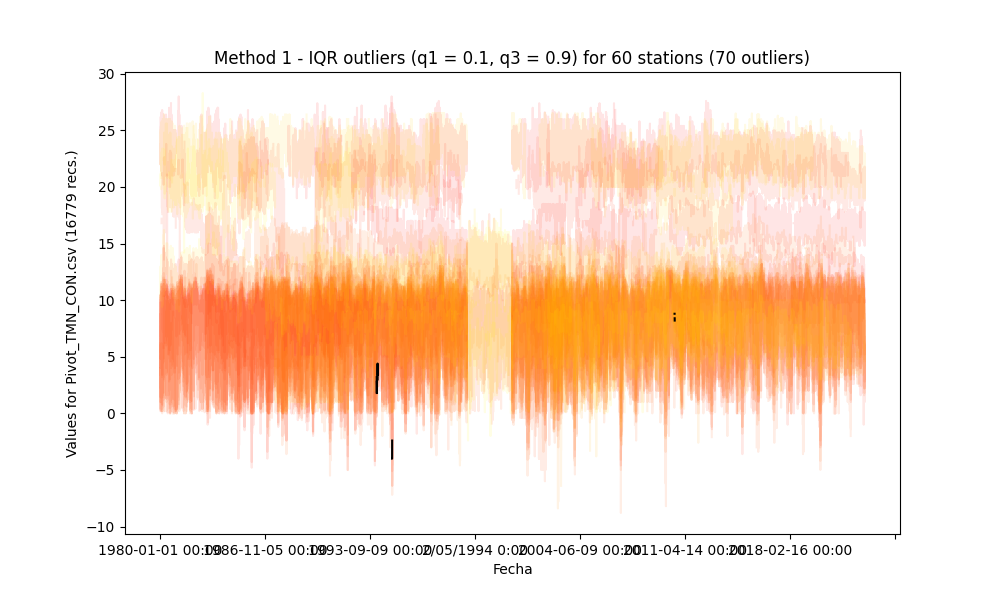
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Resultados Outlier Método 3 Temperatura Máxima



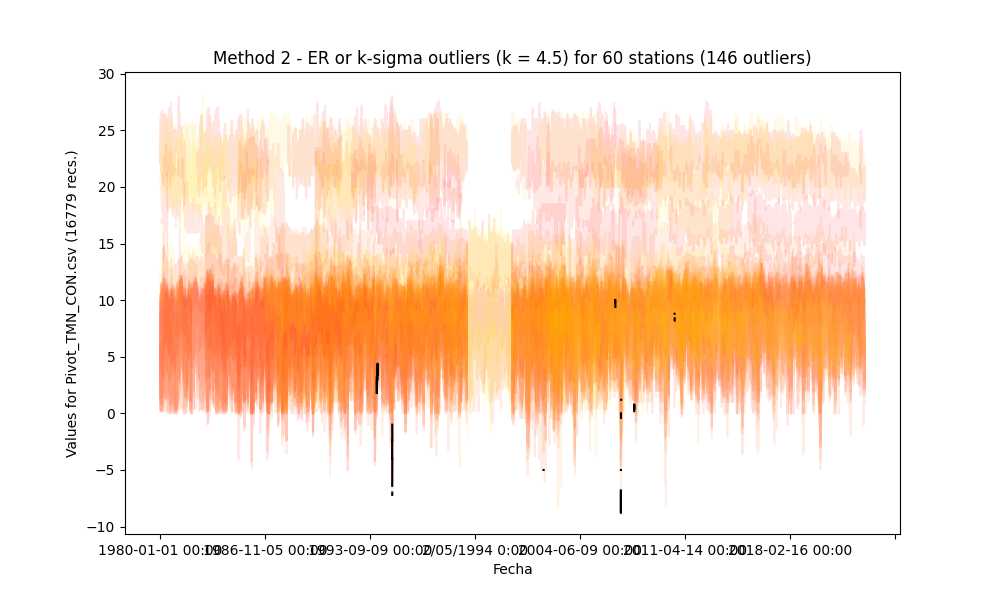
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Resultados Outlier Método 1 Temperatura Mínima



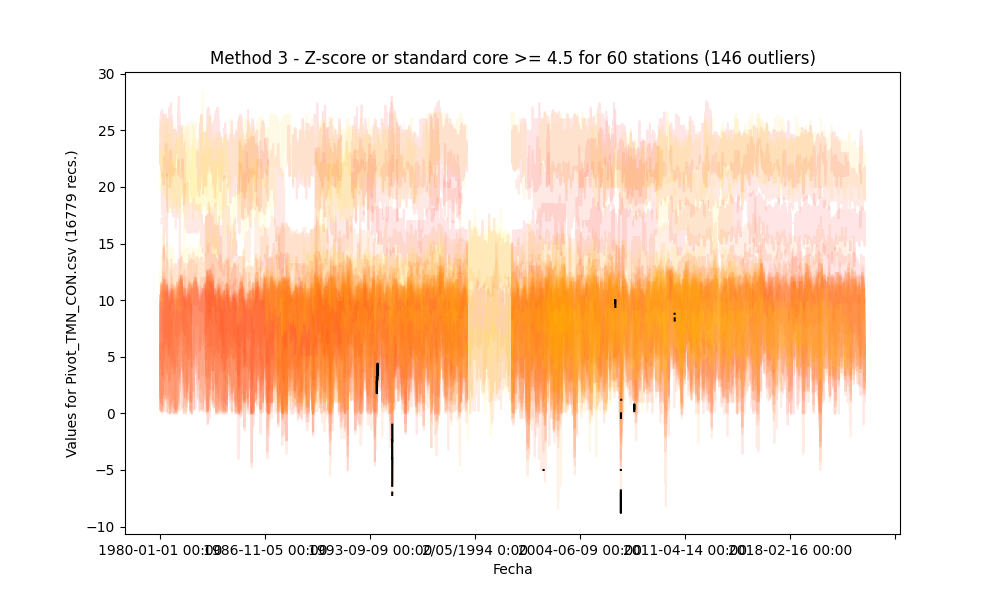
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Resultados Outlier Método 2 Temperatura Mínima



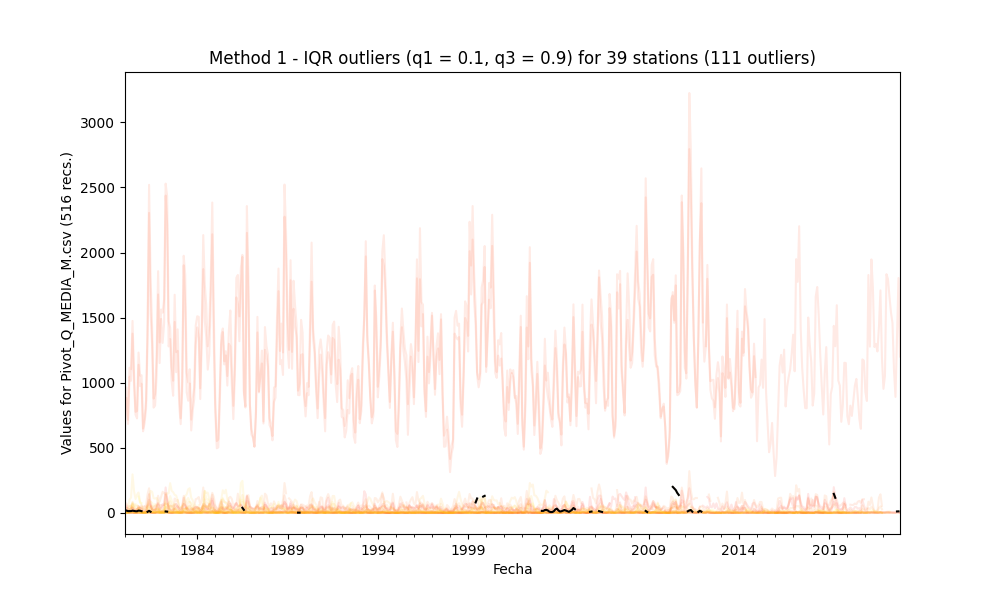
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Resultados Outlier Método 3 Temperatura Mínima



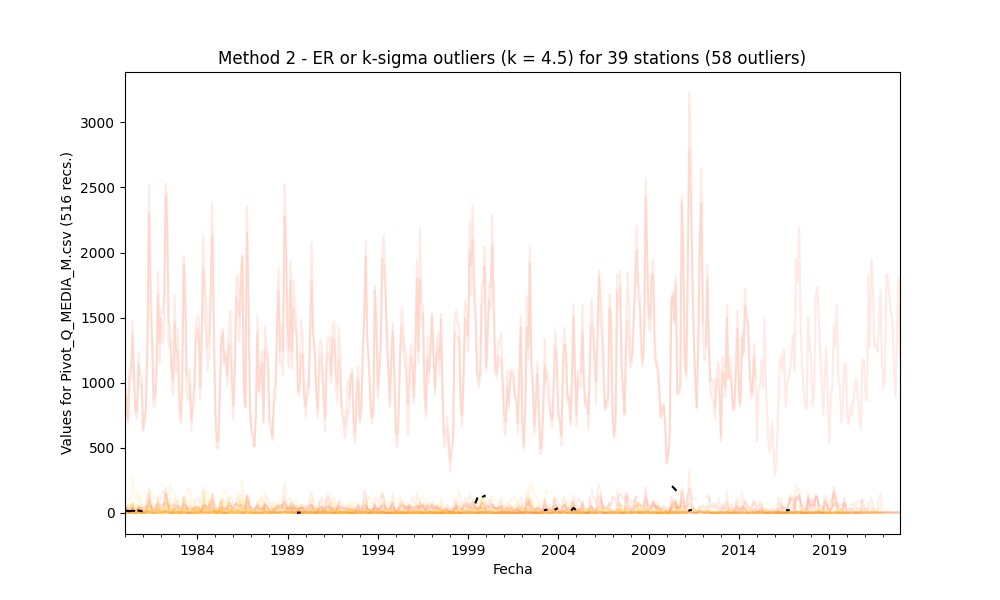
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Resultados Outlier Método 1 Caudal



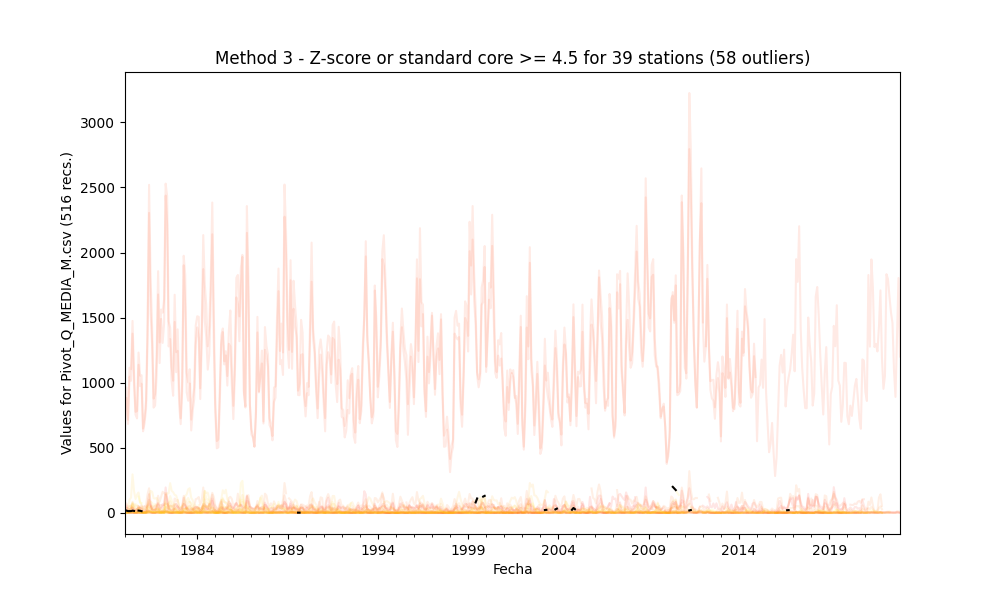
Fuente: Elaboración Propia, 2023.

Ilustración ‑. Resultados Outlier Método 2 Caudal



Fuente: Elaboración Propia, 2023.

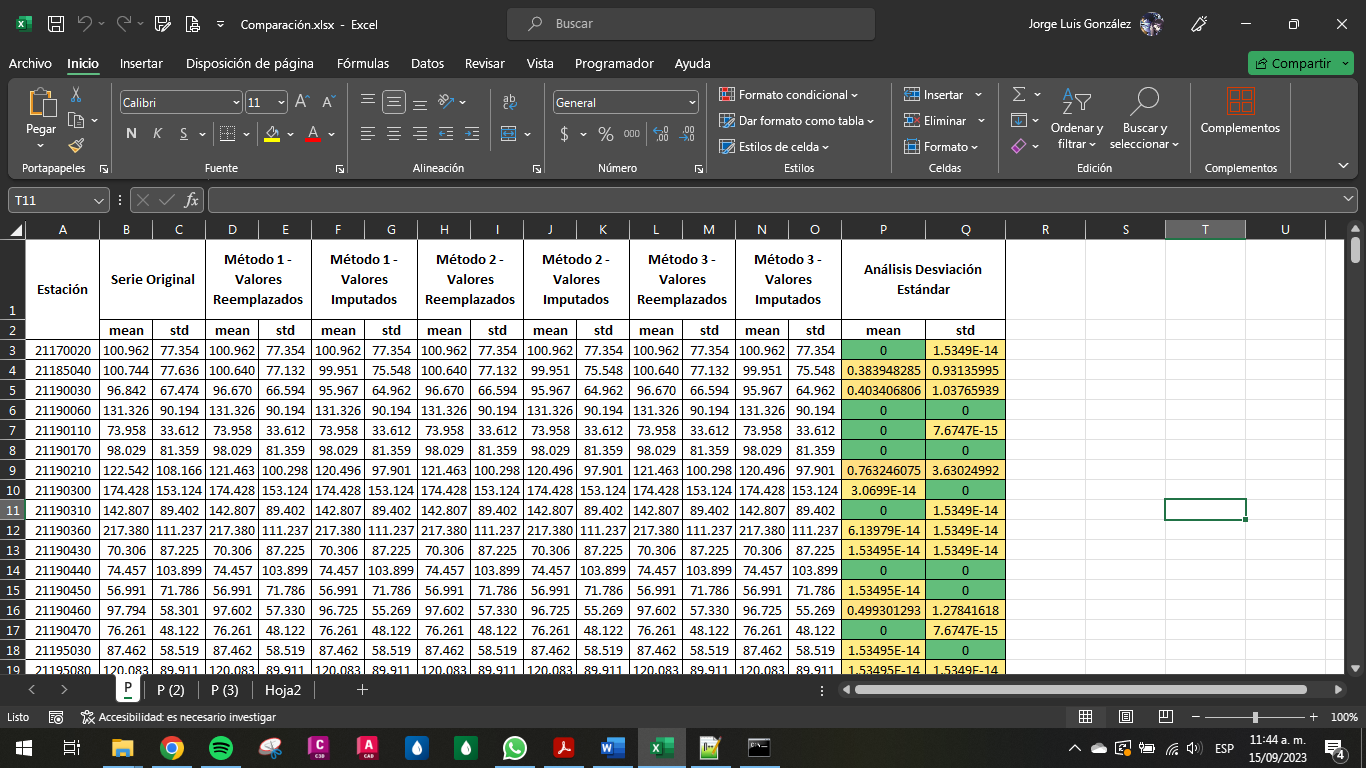
Ilustración ‑. Resultados Outlier Método 3 Caudal



Fuente: Elaboración Propia, 2023.

En la siguiente ilustración se observa el resultado de las medias y desviaciones estándar de los resultados del ejercicio del numeral anterior, donde se compara en cada estación estos parámetros para la serie original y las series con datos reemplazados e imputados por cada uno de los tres métodos usados. A estos valores estadísticos se les calculo la desviación estándar para conocer la dispersión entre sí y valorar cuales estaciones tienen grandes diferencias entre los datos iniciales y los procesados.

Ilustración ‑. Registros Precipitación

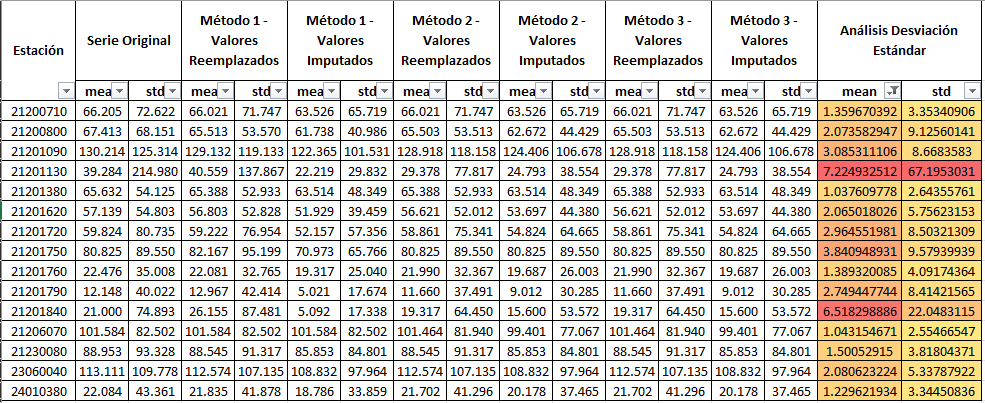


Fuente: Elaboración Propia, 2023.

Se define como el umbral de aceptación las desviaciones que sean menores a 1. Se hizo un filtro para la desviación estándar de las medias y la desviación estándar de este mismo parámetro, con el fin de identificar en cada caso la magnitud de estaciones con problemas.

Para el caso de la precipitación se obtuvieron los siguientes resultados donde se observa que el filtro por la desviación estándar genera 15 estaciones con datos deficientes que deberían ser eliminadas para análisis posteriores.

Ilustración ‑. Análisis Desviaciones Estándar Precipitación



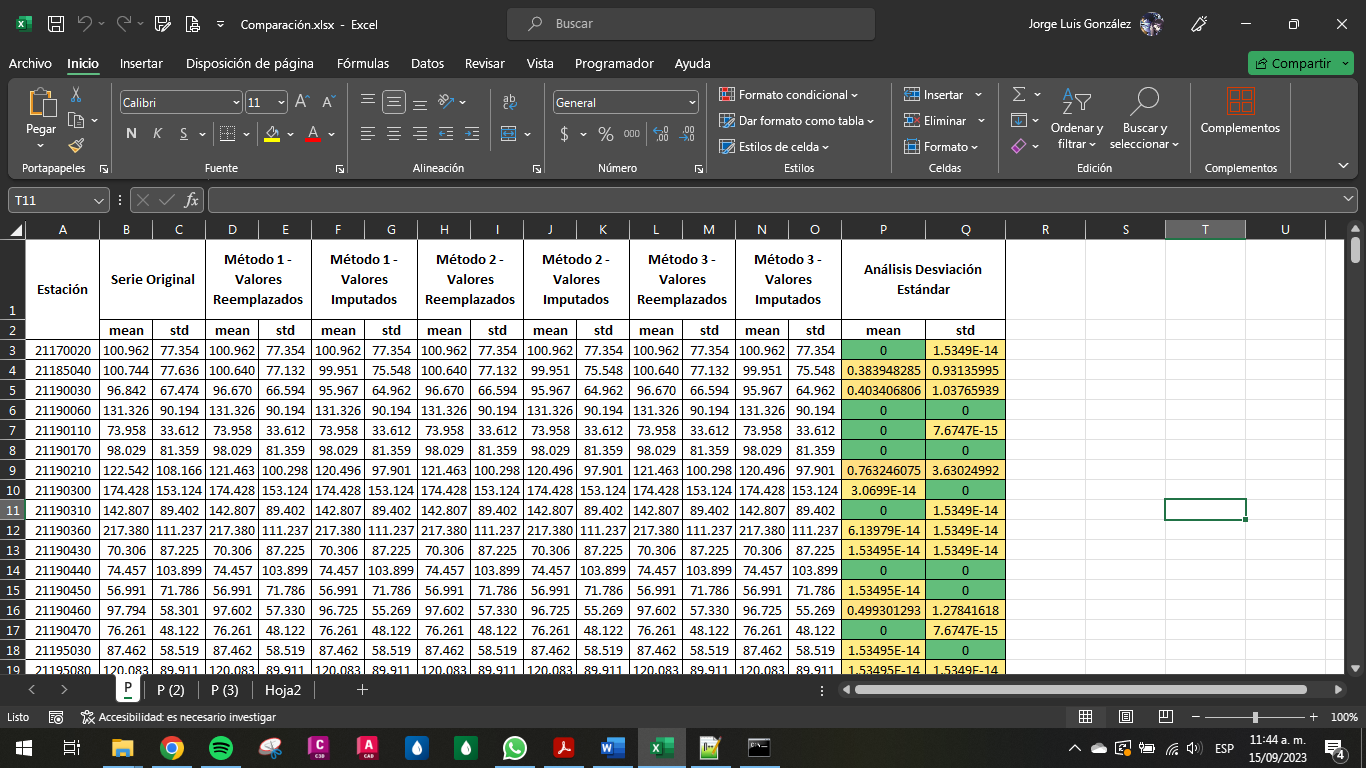
Fuente: Elaboración Propia, 2023.

Es así que las siguientes estaciones se consideran no apropiadas para continuar:

* 21200710
* 21200800
* 21201090
* 21201130
* 21201380
* 21201620
* 21201720
* 21201750
* 21201760
* 21201790
* 21201840
* 21206070
* 21230080
* 23060040
* 24010380

En la siguiente ilustración se observa el resultado de este análisis para las estaciones de evaporación.

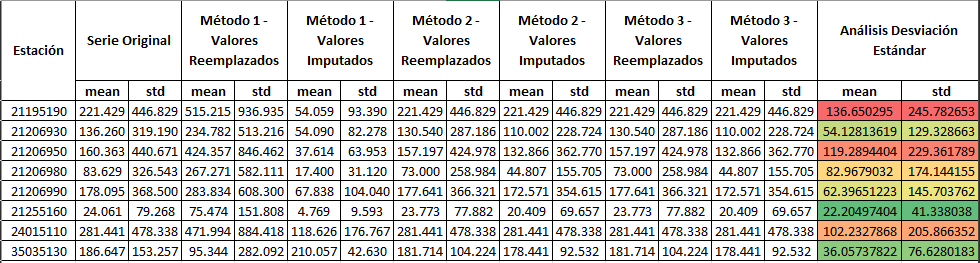
Ilustración ‑. Registros Precipitación



Fuente: Elaboración Propia, 2023.

Para el caso de la evaporación se obtuvieron los siguientes resultados donde se observa que el filtro por la desviación estándar genera 6 estaciones con datos deficientes que deberían ser eliminadas para análisis posteriores.

Ilustración ‑. Análisis Desviaciones Estándar Evaporación



Fuente: Elaboración Propia, 2023.

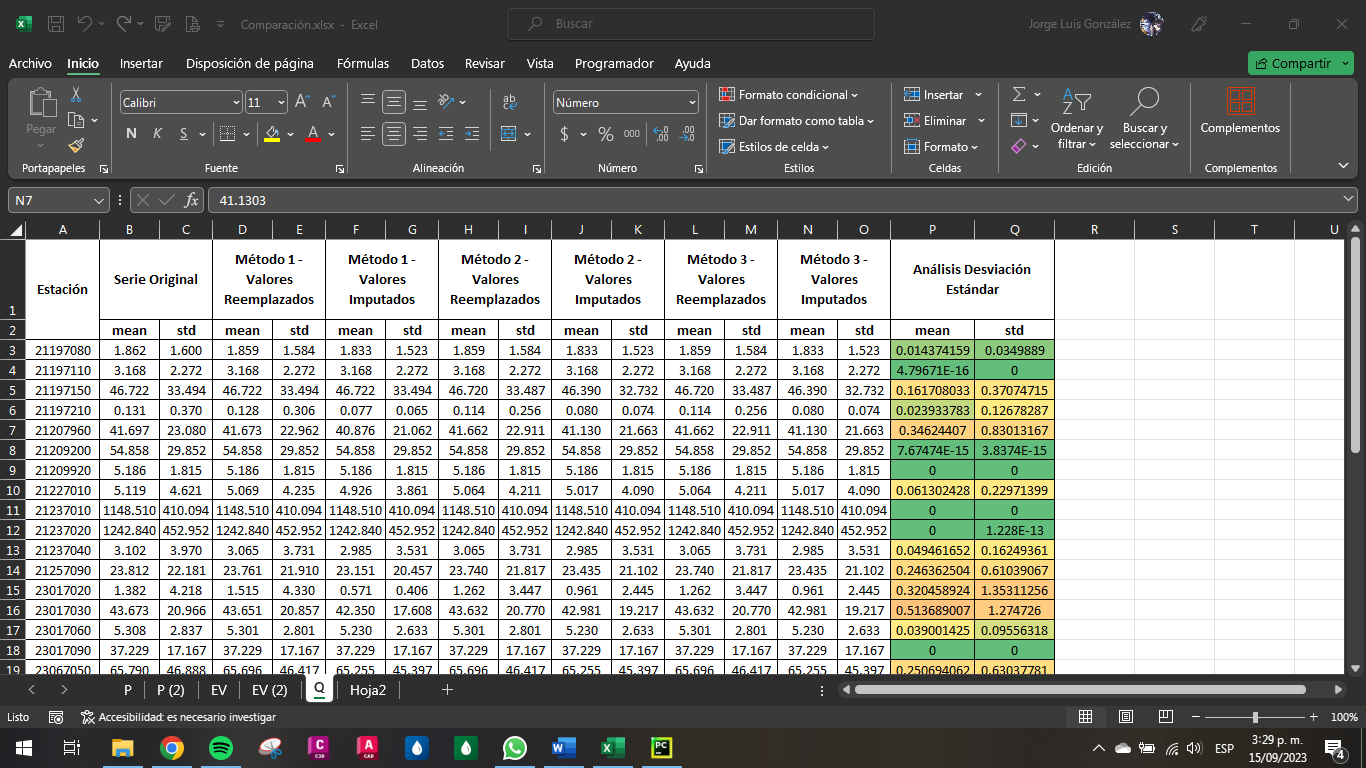
Es así que las siguientes estaciones se consideran no apropiadas para continuar:

* 21206930
* 21206950
* 21206980
* 21206990
* 21255160
* 35035130

Al ser casi la totalidad de estaciones de evaporación con problemas, podría evaluarse la sustitución de datos con otra metodología para generar las series sintéticas en cada estación o trabajar con datos de alta incertidumbre hacia el error.

En la siguiente ilustración se observa el resultado de este análisis para las estaciones de caudal.

Ilustración ‑. Registros Caudal



Fuente: Elaboración Propia, 2023.

Para el caso de caudal se obtuvieron los siguientes resultados donde se observa que el filtro por la desviación estándar genera 1 estación que es la 35067050 con datos deficientes que debería ser eliminada para análisis posteriores.

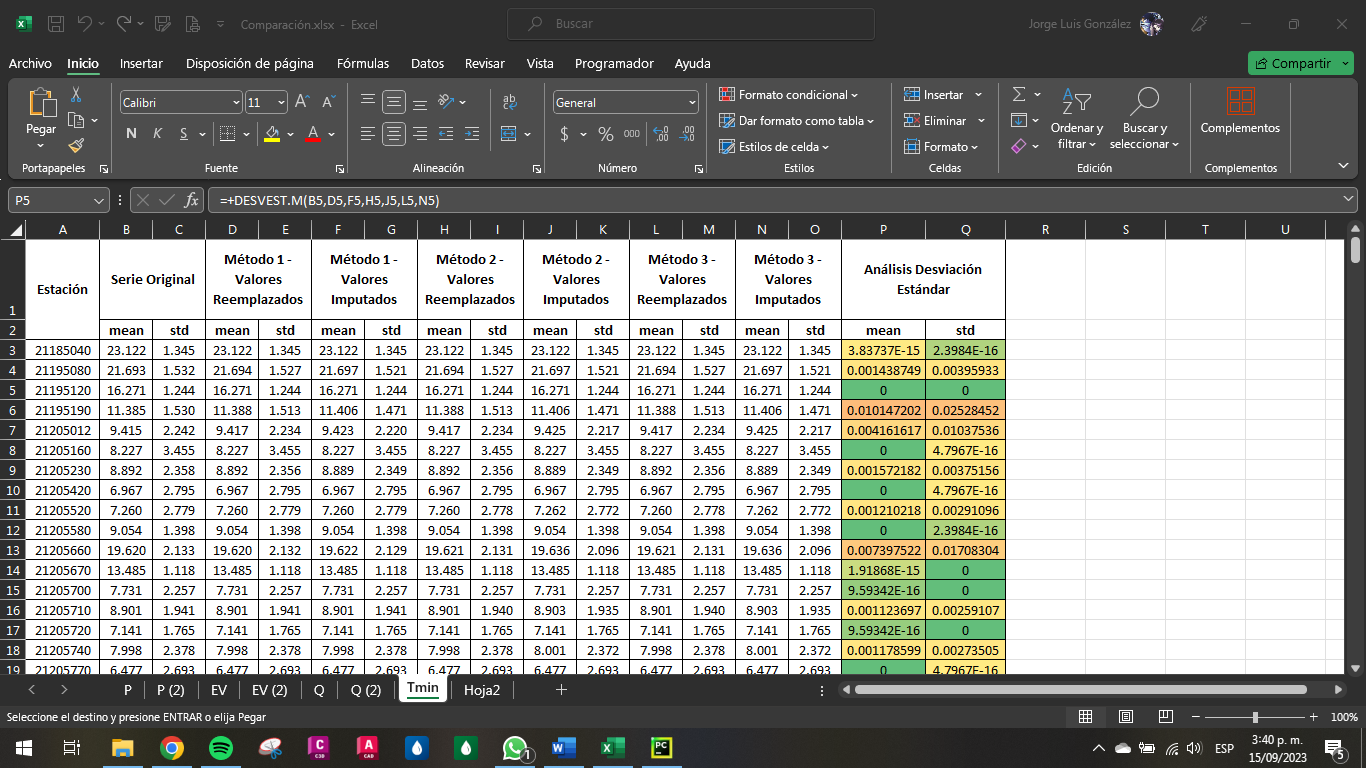
Ilustración ‑. Análisis Desviaciones Estándar Caudal



Fuente: Elaboración Propia, 2023.

En la siguiente ilustración se observa el resultado de este análisis para las estaciones de temperatura mínima.

Ilustración ‑. Registros Temperatura Mínima

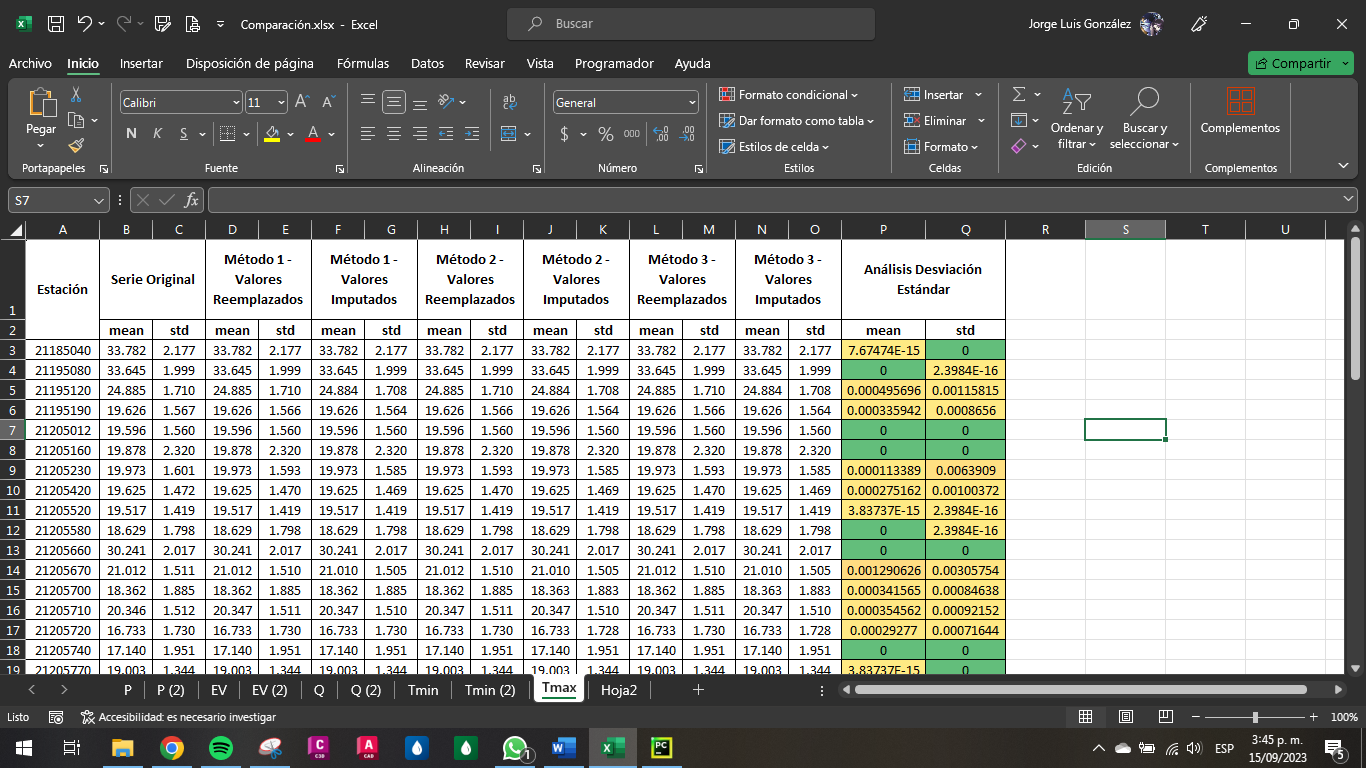


Fuente: Elaboración Propia, 2023.

Para el caso de la temperatura mínima no se obtuvieron estaciones con mala desviación estándar por lo que se podría continuar con todas.

En la siguiente ilustración se observa el resultado de este análisis para las estaciones de temperatura máxima.

Ilustración ‑. Registros Temperatura Máxima



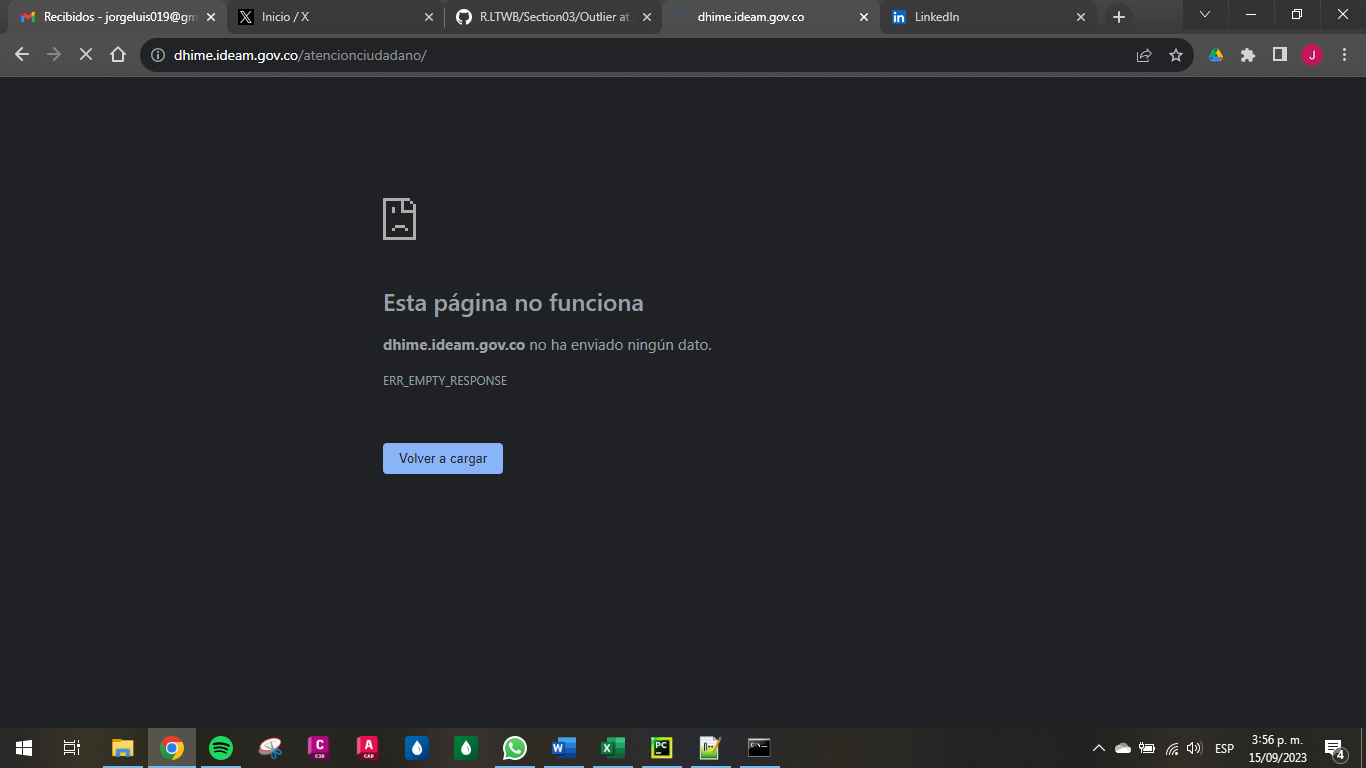
Fuente: Elaboración Propia, 2023.

Para el caso de la temperatura máxima no se obtuvieron estaciones con mala desviación estándar por lo que se podría continuar con todas.

# ACTIVIDAD 2: análisis de OTROS PARÁMETROS

A la fecha de desarrollo de esta actividad, la página web del IDEAM se encuentra deshabilitada por lo que no se pueden descargar los datos de otros parámetros.

Ilustración ‑. Acceso plataforma IDEAM



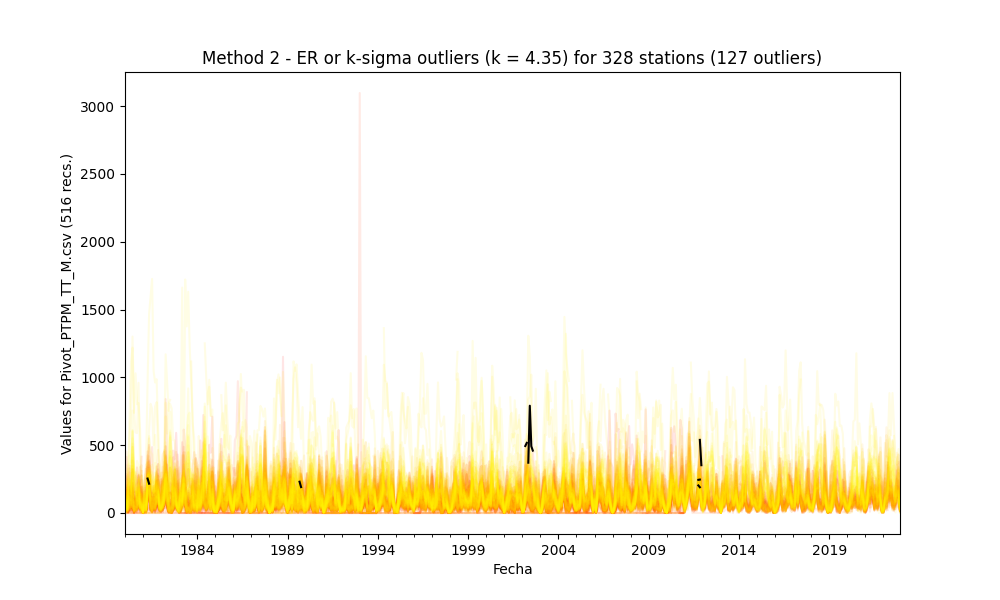
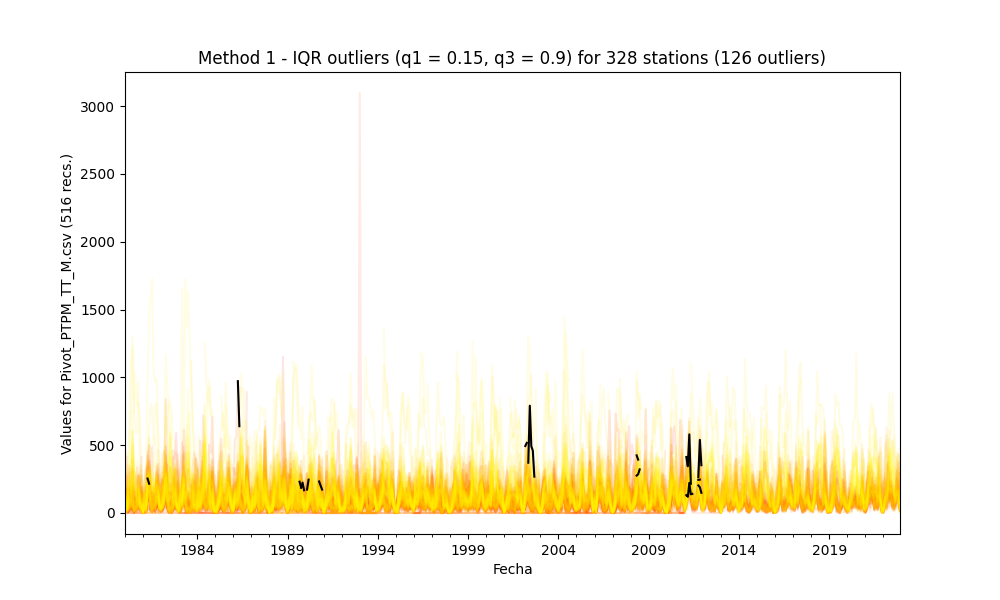
Fuente: IDEAM, 15/09/2023.

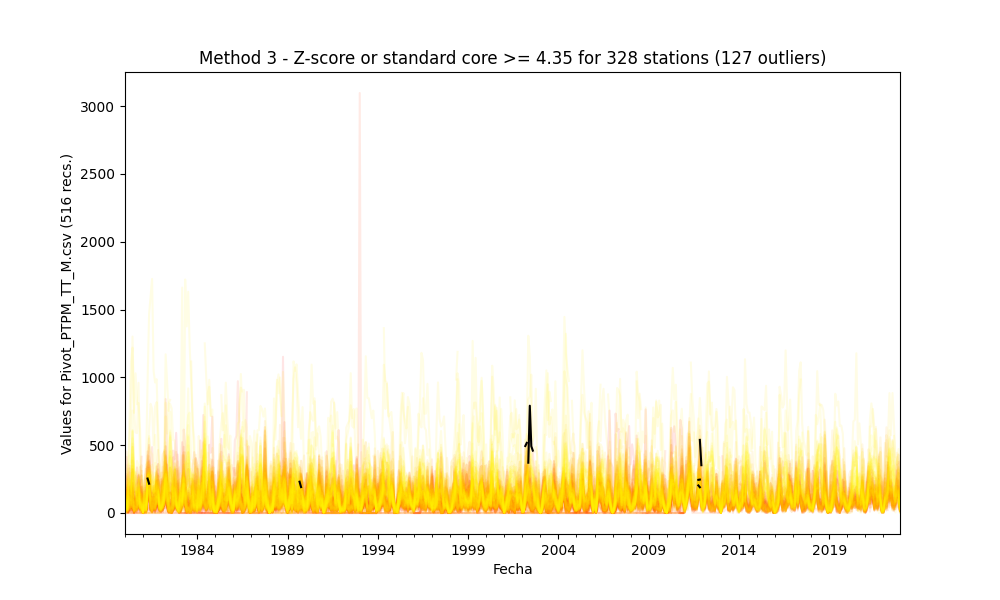
# ACTIVIDAD 3: SENSIBILIDAD PARÁMETROS ESTADÍSTICOS

Se realizó un ejercicio de ensayo y error para determinar el q1, q3, K-sigma y Z-score para generar un número similar de outliers.

Para la precipitación se tiene que q1=0.15, q3=0.90, K-sigma=4.35 y Z-score=4.35.

Ilustración ‑. Resultados Precipitación

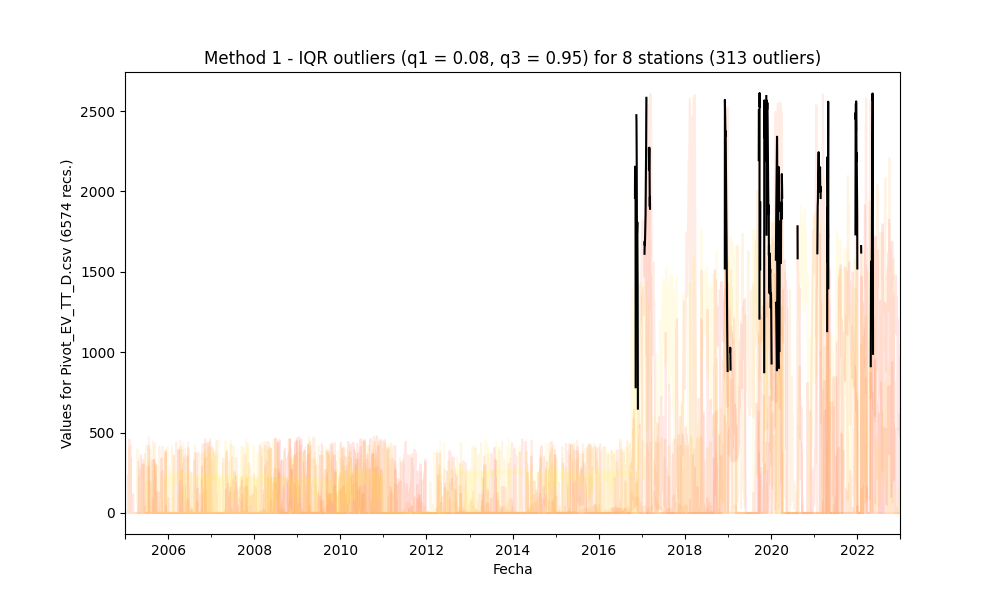


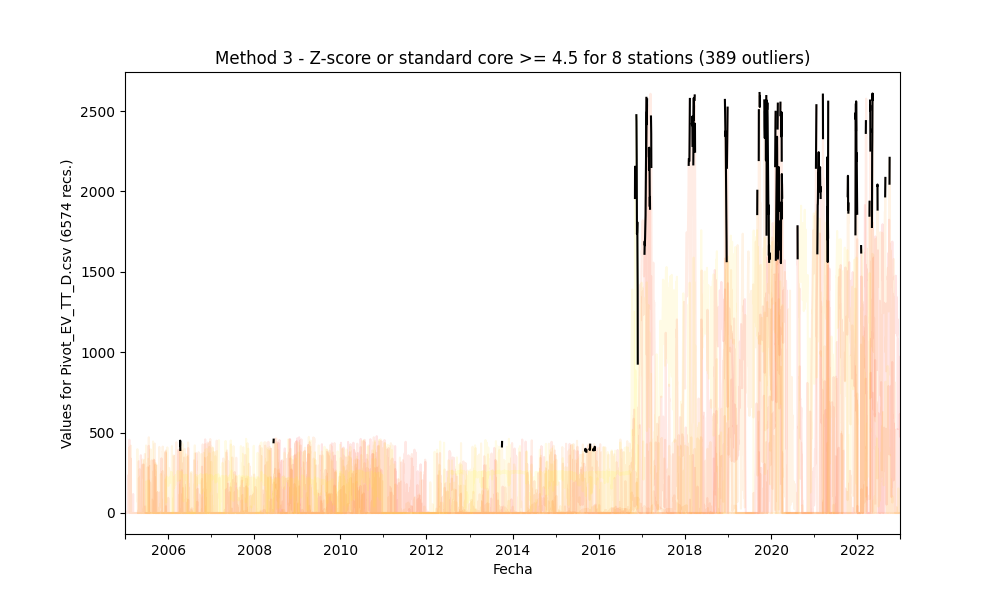
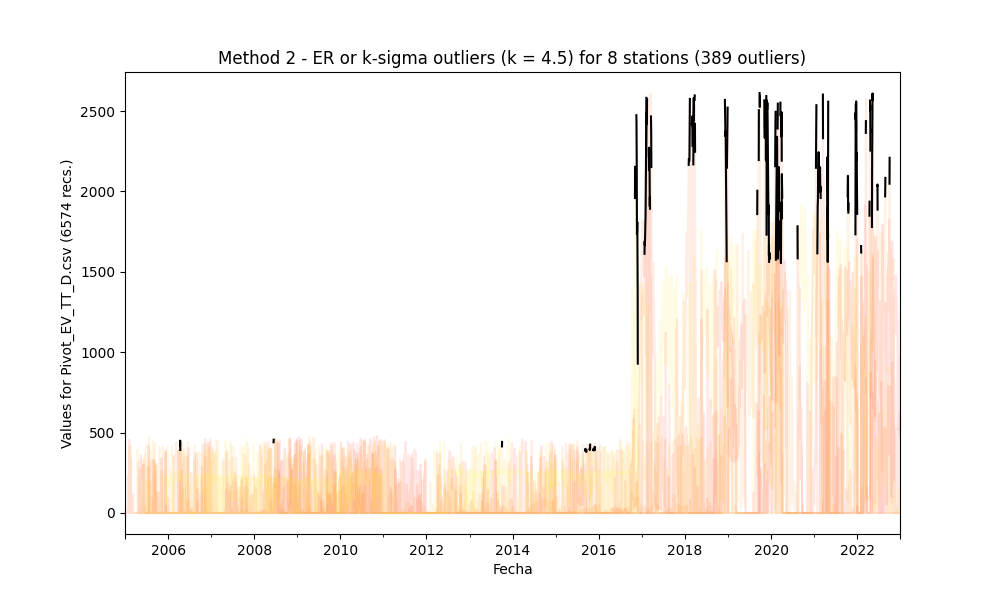


Fuente: Elaboración Propia.

Para evaporación se tiene que q1=0.08, q3=0.95, K-sigma=4.5 y Z-score=4.5.

Ilustración ‑. Resultados Evaporación

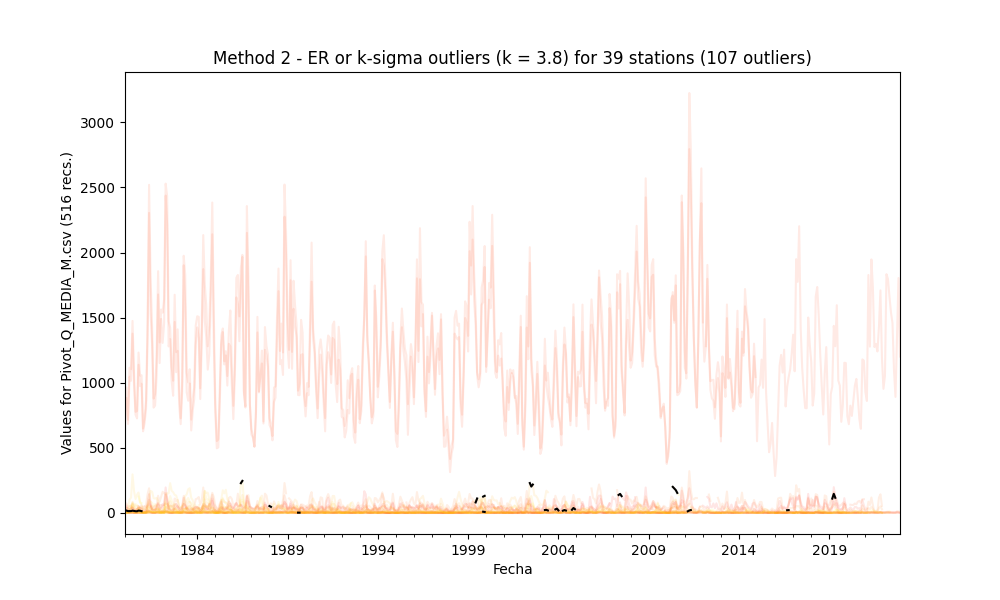
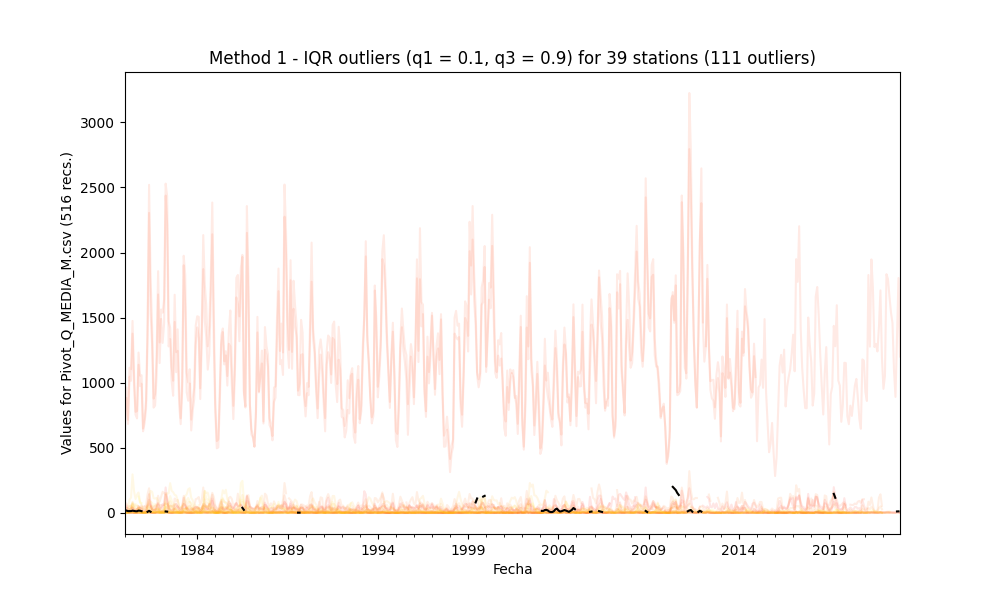


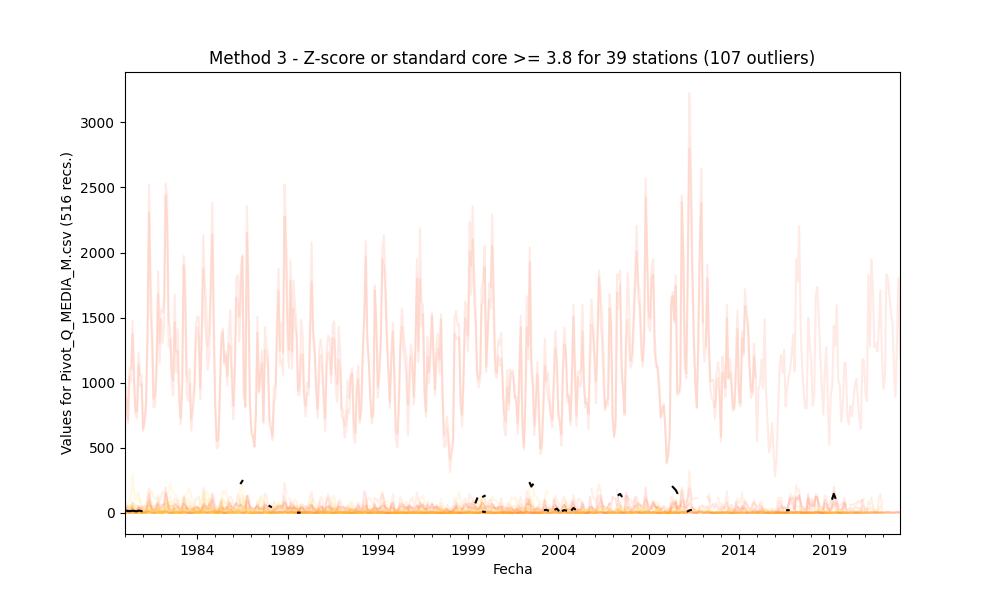


Fuente: Elaboración Propia.

Para caudal se tiene que q1=0.1, q3=0.9, K-sigma=3.8 y Z-score=3.8.

Ilustración ‑. Resultados Caudal

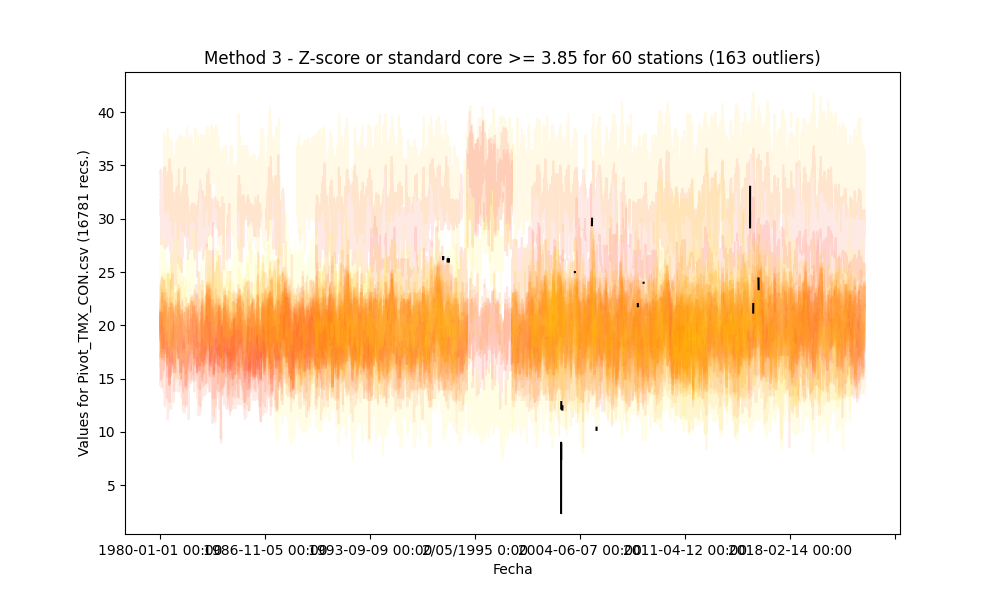


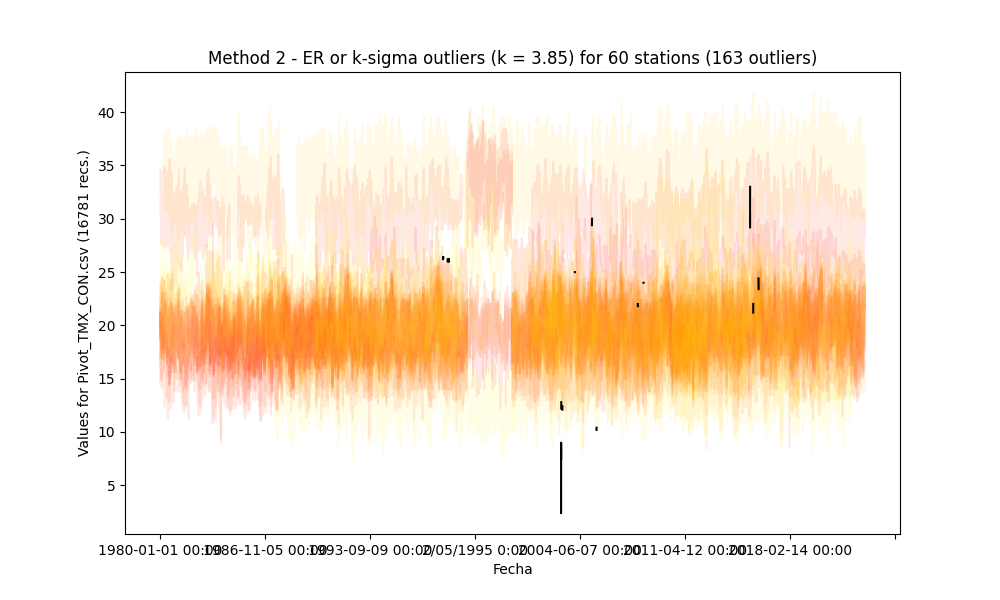
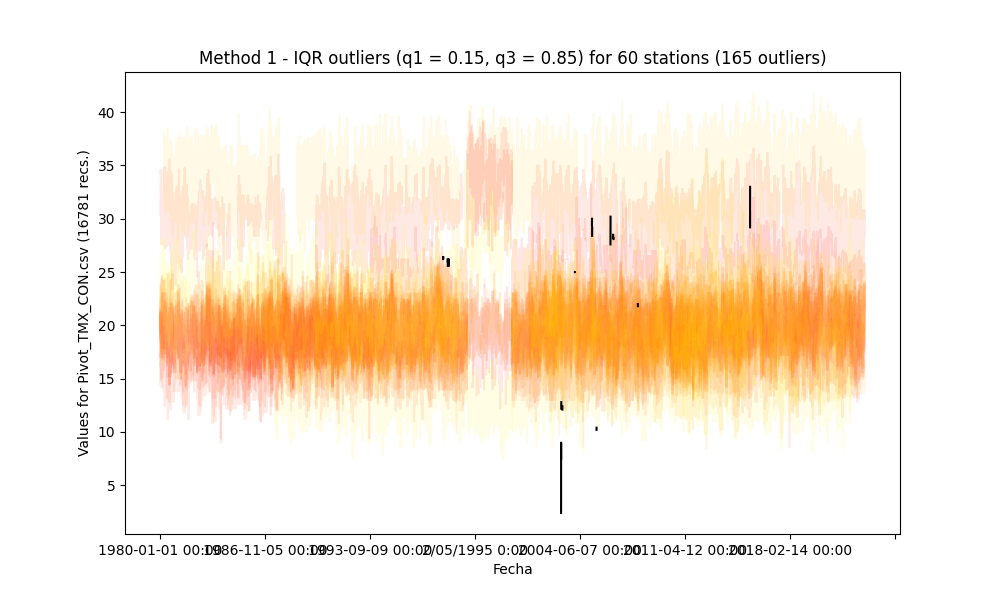


Fuente: Elaboración Propia.

Para la temperatura máxima se tiene que q1=0.15, q3=0.85, K-sigma=3.8 y Z-score=3.8.

Ilustración ‑. Resultados Temperatura Máxima

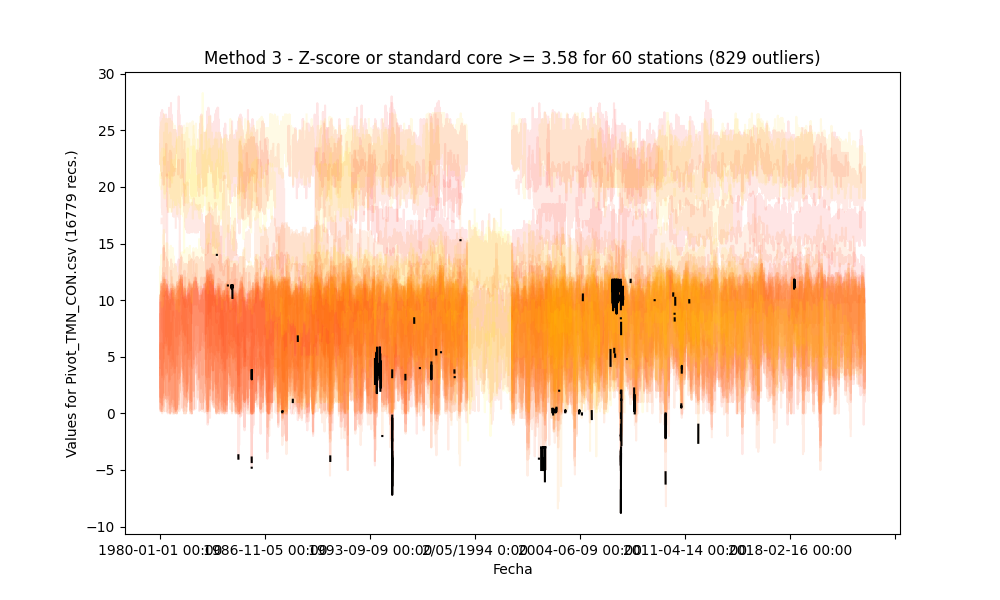
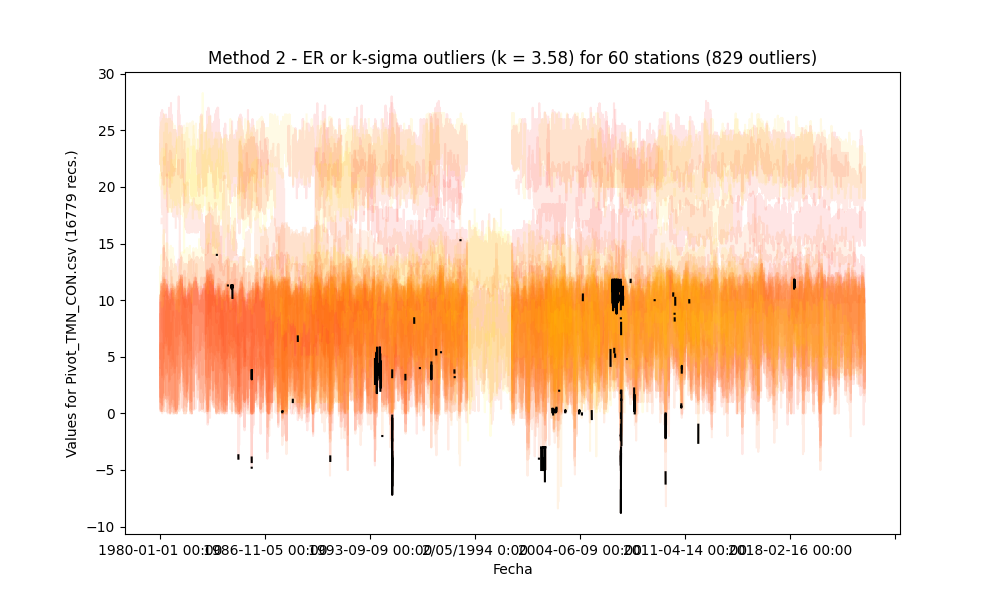
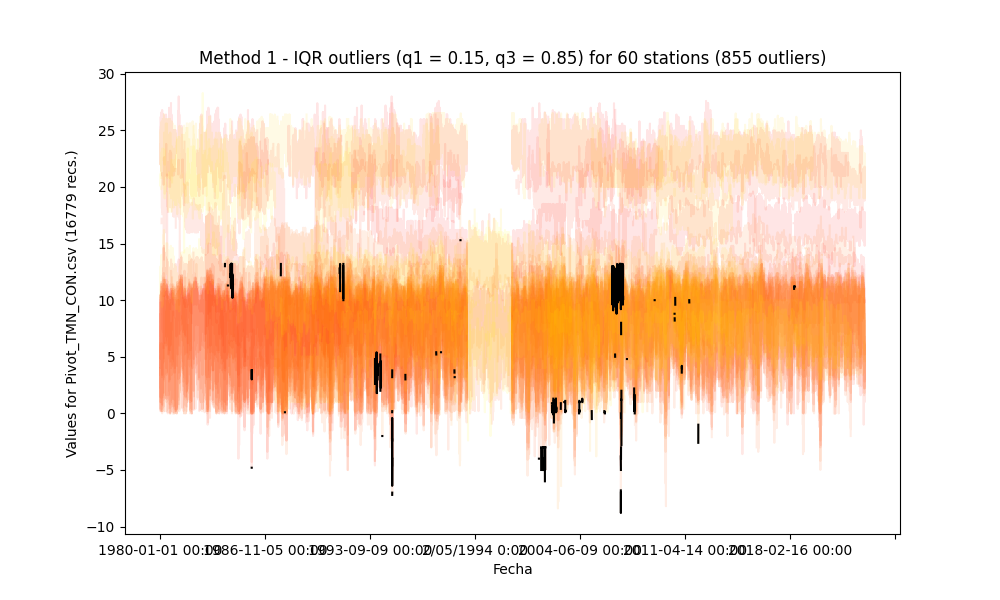




Fuente: Elaboración Propia.

Para la temperatura mínima se tiene que q1=0.15, q3=0.85, K-sigma=3.58 y Z-score=3.58.

Ilustración ‑. Resultados Temperatura Mínima



Fuente: Elaboración Propia.

# ACTIVIDAD 4: EXCLUSIÓN DE ESTACIONES

De acuerdo con los resultados obtenidos en el numeral 3 y 5 de este informe, a continuación se listan las estaciones excluidas para continuar con el desarrollo del análisis en actividades posteriores; también se tuvo en cuenta las estaciones con registros cortos de la actividad EDA.

* 21200710
* 21200800
* 21201090
* 21201130
* 21201380
* 21201620
* 21201720
* 21201750
* 21201760
* 21201790
* 21201840
* 21206070
* 21230080
* 23060040
* 24010380
* 21206100
* 35060280
* 35067050

# Conclusiones

* Se realizó el análisis de los outliers a partir de las tres metodologías incluidas en el script Outlier.py.
* Se realizó la identificación de las estaciones con mejor ajuste de datos reemplazados o imputados.
* Se realizó el análisis de sensibilidad de los parámetros el q1, q3, K-sigma y Z-score.
* No se pude realizar la actividad con otros parámetros ya que a la fecha la plataforma de descarga del IDEAM no se encuentra activa.
* Se definieron las estaciones a excluir para análisis posteriores.

# Referencias Bibliográficas

* RCFDTOOLS, 2023. Balance hidrológico de largo plazo para estimación de caudales medios usando SIG. Contenido del curso: https://github.com/rcfdtools/R.LTWB/tree/main/Section03/Outlier.