



## R.LTWB – SECTION 03

Descarga, procesamiento y análisis de datos  
hidroclimatológicos  
identificación y procesamiento de datos  
atípicos - Outliers

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

## TABLA DE CONTENIDO

1. Introducción .....	3
2. Objetivo General .....	3
3. Actividad 1: Procesamiento en software .....	3
4. ACTIVIDAD 2: análisis de OTROS PARÁMETROS .....	3
5. ACTIVIDAD 3: SENSIBILIDAD PARÁMETROS ESTADÍSTICOS .....	3
6. Conclusiones.....	11
7. Referencias Bibliográficas .....	1

## ÍNDICE DE ILUSTRACIONES

Ilustración 3-1. Descarga Outlier.py .....	3
Ilustración 3-2. Script Outlier.py .....	4
Ilustración 3-3. Ejecución inicial Outlier.py .....	4
Ilustración 3-4. Resultados Precipitación script Outlier.py .....	6
Ilustración 3-5. Resultados Outlier Método 1 Precipitación .....	7
Ilustración 3-6. Resultados Outlier Método 2 Precipitación .....	7
Ilustración 3-7. Resultados Outlier Método 3 Precipitación .....	8
Ilustración 3-8. Resultados Outlier Método 1 Evaporación .....	8
Ilustración 3-9. Resultados Outlier Método 2 Evaporación .....	9
Ilustración 3-10. Resultados Outlier Método 3 Evaporación .....	9
Ilustración 3-11. Resultados Outlier Método 1 Temperatura Máxima .....	10
Ilustración 3-12. Resultados Outlier Método 2 Temperatura Máxima .....	10
Ilustración 3-13. Resultados Outlier Método 3 Temperatura Máxima .....	11
Ilustración 3-14. Resultados Outlier Método 1 Temperatura Mínima .....	11
Ilustración 3-15. Resultados Outlier Método 2 Temperatura Mínima .....	12
Ilustración 3-16. Resultados Outlier Método 3 Temperatura Mínima .....	12
Ilustración 3-17. Resultados Outlier Método 1 Caudal .....	13
Ilustración 3-18. Resultados Outlier Método 2 Caudal .....	13
Ilustración 3-19. Resultados Outlier Método 3 Caudal .....	14
Ilustración 3-20. Registros Evaporación .....	14
Ilustración 3-21. Análisis Desviaciones Estándar Precipitación .....	15
Ilustración 3-22. Registros Evaporación .....	1
Ilustración 3-23. Análisis Desviaciones Estándar Evaporación .....	1
Ilustración 3-24. Registros Caudal .....	1
Ilustración 3-25. Análisis Desviaciones Estándar Caudal .....	1
Ilustración 3-26. Registros Temperatura Mínima .....	2
Ilustración 3-27. Registros Temperatura Máxima .....	2
Ilustración 4-1. Acceso plataforma IDEAM .....	3
Ilustración 5-1. Resultados Precipitación .....	4

SECTION 03  
DESCARGA, PROCESAMIENTO  
Y ANÁLISIS DE DATOS  
HIDROCLIMATOLÓGICOS

JORGE LUIS GONZÁLEZ CASTRO  
CC: 1032395475  
CS2021

Ilustración 5-2. Resultados Evaporación .....	5
Ilustración 5-3. Resultados Caudal .....	7
Ilustración 5-4. Resultados Temperatura Máxima .....	8
Ilustración 5-5. Resultados Temperatura Mínima .....	9

## 1. 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/ilgingcivil/R.LTWB.CS2021> para la inclusión de los archivos y documentos de las actividades desarrolladas.

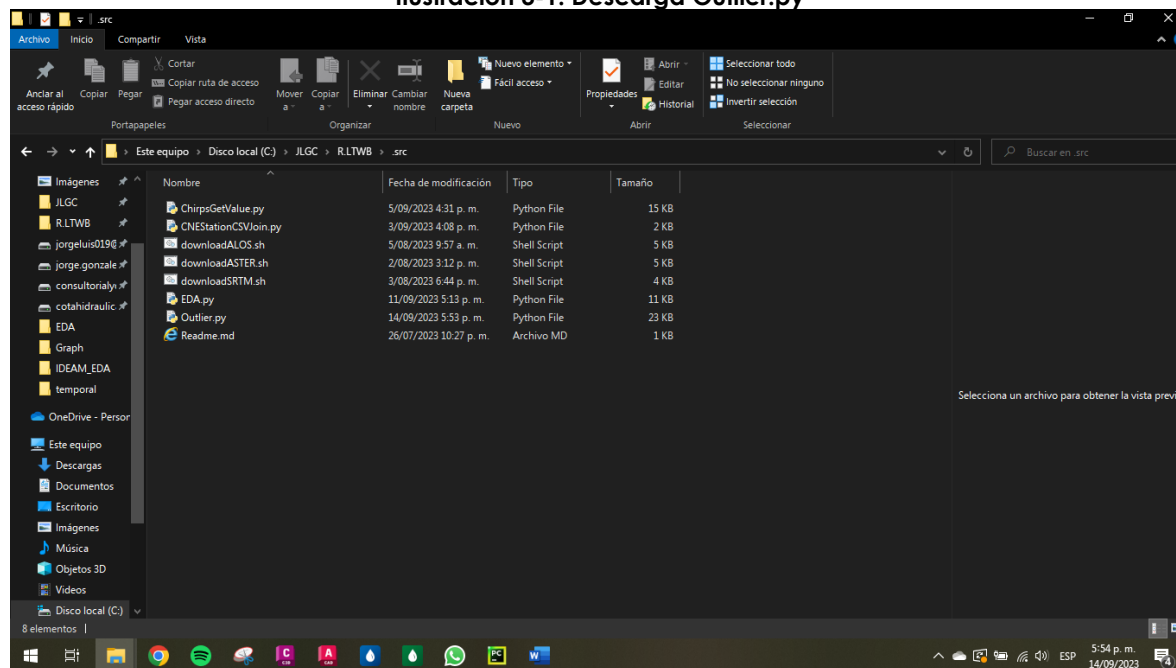
## 2. OBJETIVO GENERAL

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.

## 3. 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 3-1. 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 3-2. Script Outlier.py

```

133 z = (df - df.mean()) / df.std()
134 if drop:
135     outliers = df[~(abs(z) >= zscore_threshold)]
136 else:
137     outliers = df[(abs(z) >= zscore_threshold)]
138 return outliers
139
140
141 # General variables
142 pivot_table_name = 'Pivot_PTPM_TT_M.csv' # <<<<< Pivot table name to process
143 path_input = 'C:/JLGC/R.LTWB/datasets/IDEAM_EDA/' # Current location from pivot tables
144 station_file = path_input + pivot_table_name # Current pivot IDEAM records file for a specified parameter
145 path = 'C:/JLGC/R.LTWB/datasets/IDEAM_Outlier/' # Your local output path, use ../datasets/IDEAM_Outlier/ for relative path
146 file_log_name = path + 'Outlier_IQR_' + pivot_table_name + '.md' # Markdown file log
147 file_log = open(file_log_name, 'w+') # w+ create the file if it doesn't exist
148 date_record_name = 'Fecha' # IDEAM date field name for the record values
149 plot_colormap = 'autumn' # Color theme for plot graphics, https://matplotlib.org/stable/tutorials/colors/colormaps.html
150 sample_records = 3 # Records to show in the sample table head and tail
151 fig_size = 5 # Height size for figures plot
152 print_table_sample = True
153 show_plot = True
154 station_exclude = ['28017140', '25027020', '25027410', '25027490', '25027330', '25027390', '25027630', '25027360', '25027320', '16067010', '25027420'] # Use [
155 ql_val = 0.1 # Default is 0.25. Method 1.
156 q3_val = 0.9 # Default is 0.75. Method 1.
157 cap_multiplier = 4.5 # Replacement cap outlier value multiplier or k value, default is 3. e.j, mean() +/- cap_multiplier * std(). k over empirical rules. . Met
158 zscore_threshold = 4.5 # Z-score threshold, default is 3. Method 3. If the threshold is equal to the cap_multiplier, the results are the same as Method 2.
159
160
161 # Header
162 print_log('# Outliers detection and processing through statistical methods')
163 print_log('\n Processed file: [%s] (%s)' % (str(station_file), '..\IDEAM_EDA\' + pivot_table_name) +
164 '\n Execution date: ' + str(datetime.now()) +
165 '\n Python version: ' + str(sys.version) +
166 '\n Python path: ' + str(sys.path[0:5]) +
167 '\n matplotlib version: ' + str(matplotlib.__version__) +
168 '\n')
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
```

SECTION 03  
DESCARGA, PROCESAMIENTO  
Y ANÁLISIS DE DATOS  
HIDROCLIMATOLÓGICOS

JORGE LUIS GONZÁLEZ CASTRO  
CC: 1032395475  
CS2021

Simbolo del sistema

49	23	18	14	36	15	20	8	1	8	5	44	44
1	492	26	48	78								

General statistics table - Initial file

	count	mean	std	min	25%	50%	75%	max
21170020	458	100.962	77.3544	0	40.525	81.95	148.225	446
21185040	507	100.744	77.6364	0	39.55	87.6	143.75	502.7
21190030	432	96.8424	67.4744	0.4	44.125	87.2	128.75	475.1
21190060	156	131.326	90.1943	0	61.5	132.5	201	342
21190110	12	73.9583	33.6117	7	53.75	77.5	97.25	118.5
21190170	7	98.0286	81.3592	14.4	69.95	78.1	90.75	272.3
21190210	503	121.463	100.298	0	43.15	101.9	177	609.291
21190300	376	174.428	153.124	0	65	135.4	237.725	764
21190310	499	142.807	89.402	6.6	70.15	124	196.45	479
21190360	5	217.38	111.237	101.9	102	248	287	348
21190430	32	70.3063	87.225	0	0	43.45	100.8	351.4
21190440	30	74.4567	103.899	0	0	22.05	142.35	413.9
21190450	32	56.9906	71.7864	0	0	29.4	92.2	250.8
21190460	299	97.602	57.3296	0	57.95	85.5	133.05	360.147
21190470	156	76.2609	48.1221	0	41.725	71.75	106.45	205.3
21195030	180	87.4622	58.519	5	37.1	80	124.3	283.9
21195080	236	120.083	89.9106	0	44.35	107.95	172.175	385.9
21195120	472	115.896	70.8954	1.9	60.55	105.65	158.725	352.2
21195140	174	56.4833	61.441	0	0	38.95	89.15	277.2
21195190	506	74.6785	46.4629	0	38.75	64.25	104.725	262.6
21200040	12	30.35	14.5619	0.7	20.825	33.4	36.15	55.8
21200080	354	106.725	60.8913	4.6	62.225	97.65	135.7	378.3
21200110	355	65.4727	40.51	1.7	36.8	57.7	84.75	261.7
21200130	373	105.404	60.1798	1	62.7	99.5	133.7	410.9
21200160	443	94.2102	63.1426	0	46.6	85.7	130.75	369.5
21200170	11	61.4727	39.7307	19.6	26.25	56.8	78.85	130.3
21200190	330	118.996	72.3121	0	66.9	111.75	157.95	345.2
21200200	376	68.6545	41.1936	0.1	39.525	62.8	89.875	230.1
21200230	305	89.9007	68.4203	1.4	38.0	73.5	100.2	345.2
21200240	377	110.255	59.6102	0	69.7	104.2	145.4	372.6
21200260	338	84.3908	55.9034	0	46.475	73.1	114.15	415.5
21200270	304	64.0273	46.3141	0	28.575	54.1	91.85	253
21200310	361	77.3424	53.1054	0.9	34.8	66.2	106.6	277.6
21200320	363	96.7683	55.6073	8.3	57	85.6	128.2	369.9
21200340	375	83.384	48.569	0.9	45.15	80.7	113.25	235.1
21200380	187	106.0417	40.2679	0	0	0	0	171.2
21200390	12	77.6667	54.3814	21	34.5	55	115.75	185
21200400	364	91.825	60.0725	5.4	47.65	80.7	118.975	341.6
21200430	288	54.7118	37.0766	0	26.075	51.3	77	204.1

[R.LTWB](Outlier\_IQR\_Pivot\_PTPM\_TT\_M.csv.png)

#### Identified and cleaning tables for 117 IQR outliers founded

- \* Outliers identified file: [Outlier\_IQR\_Pivot\_PTPM\_TT\_M.csv](../datasets/IDEAM\_Outlier/Outlier\_IQR\_Pivot\_PTPM\_TT\_M.csv)
- \* Outliers dropped file: [Outlier\_IQR\_Drop\_Pivot\_PTPM\_TT\_M.csv](../datasets/IDEAM\_Outlier/Outlier\_IQR\_Drop\_Pivot\_PTPM\_TT\_M.csv)
- \* Outliers capped file: [Outlier\_IQR\_Cap\_Pivot\_PTPM\_TT\_M.csv](../datasets/IDEAM\_Outlier/Outlier\_IQR\_Cap\_Pivot\_PTPM\_TT\_M.csv)
- \* Outliers imputed file: [Outlier\_IQR\_Impute\_Pivot\_PTPM\_TT\_M.csv](../datasets/IDEAM\_Outlier/Outlier\_IQR\_Impute\_Pivot\_PTPM\_TT\_M.csv)

#### Statistical values for the capped and imputed file

IQR - General statistics table - Capped file

	count	mean	std	min	25%	50%	75%	max
21170020	458	100.962	77.3544	0	40.525	81.95	148.225	446
21185040	507	100.64	77.1318	0	39.55	87.6	143.75	450.108
21190030	432	96.6696	66.5936	0.4	44.125	87.2	128.75	400.477
21190060	156	131.326	90.1943	0	61.5	132.5	201	342
21190110	12	73.9583	33.6117	7	53.75	77.5	97.25	118.5
21190170	7	98.0286	81.3592	14.4	69.95	78.1	90.75	272.3
21190210	503	121.463	100.298	0	43.15	101.9	177	609.291
21190300	376	174.428	153.124	0	65	135.4	237.725	764
21190310	499	142.807	89.402	6.6	70.15	124	196.45	479
21190360	5	217.38	111.237	101.9	102	248	287	348
21190430	32	70.3063	87.225	0	0	43.45	100.8	351.4
21190440	30	74.4567	103.899	0	0	22.05	142.35	413.9
21190450	32	56.9906	71.7864	0	0	29.4	92.2	250.8
21190460	299	97.602	57.3296	0	57.95	85.5	133.05	360.147
21190470	156	76.2609	48.1221	0	41.725	71.75	106.45	205.3
21195030	180	87.4622	58.519	5	37.1	80	124.3	283.9
21195080	236	120.083	89.9106	0	44.35	107.95	172.175	385.9
21195120	472	115.896	70.8954	1.9	60.55	105.65	158.725	352.2
21195140	174	56.4833	61.441	0	0	38.95	89.15	277.2
21195190	506	74.6785	46.4629	0	38.75	64.25	104.725	262.6
21200040	12	30.35	14.5619	0.7	20.825	33.4	36.15	55.8
21200080	354	106.725	60.8913	4.6	62.225	97.65	135.7	378.3
21200110	355	65.4727	40.51	1.7	36.8	57.7	84.75	261.7
21200130	373	105.311	59.7315	1	62.7	99.5	133.7	376.213

Simbolo del sistema

35030230	276	177.992	107.508	2.5	95.1	167.1	247.425	522.2
35030250	229	174.745	135.803	0	55	162.9	261	599.9
35030260	377	88.779	61.2329	0	44.1	76.8	115.8	341.6
35035040	117	138.173	81.9228	7.3	76.2	136.8	181.8	382.6
35035050	72	205.138	122.56	5.2	116.825	200.1	266.15	593.4
35040010	473	552.784	333.447	0	297.3	561	766	1726
35060010	2	55.5	33.234	32	43.75	55.5	67.25	79
35060020	470	114.552	82.4426	0	54.9	99.1	163.65	477.7
35060050	506	103.263	69.7326	0	48.725	93.25	144.325	378
35060090	467	184.838	114.381	0	93.35	173	261.55	550
35060100	493	176.995	108.837	0	89.4	171.2	249.7	514.3
35060120	408	245.221	144.155	0.7	135	233.35	349.85	700
35060130	502	270.237	168.472	0	137.25	258.35	382.375	974
35060150	480	274.399	176.177	0	140.9	255.9	379.25	1051.4
35060160	501	144.727	85.8664	2	76.9	135.4	194.4	421.8
35060170	496	143.543	97.9168	0	63.475	136.95	216.6	534
35060180	508	167.54	108.167	0	83.85	154.7	235.125	573
35060200	515	149.645	93.3251	0.8	83	133.1	202.2	493.6
35060210	508	161.211	98.614	0.1	84.8	151.45	229.125	545
35060230	511	170.884	114.975	0	89.5	167.5	252	607
35060240	472	107.989	77.2206	0	46	97	152	435
35065010	472	106.152	60.82	0	61.45	101.7	149	320
35070110	515	100.079	65.6815	0	42.1	94.5	150.05	296.6
35070160	24	159.617	73.8321	55.8	105	147	179.5	359
35070230	490	94.9976	60.9017	0	45.55	88.75	134.45	310.4
35070480	468	94.9004	60.0422	0.3	46.7	90.85	134.75	279.4
35070490	438	189.588	124.709	0	88.225	174.85	268.15	603.7

```

> The _drop files_ contains the database values without the outliers identified.
>
> The _capped files_ contains the database values and the outliers has been replaced with the lower or upper capped value calculated. Lower outliers could be replaced with negative values because the limit is defined with (mean() - cap_multiplier * std()). In some cases like _temperature analysis_, the upper outliers values could be replaced with values over the original values and you can try to fix this issue changing the parameter _cap_multiplier_ that defines the stripe values range.
>
> The imputation method replace each outlier value with the mean value that contains the original outliers values.

[*1]: Adapted from: https://careerfoundry.com/en/blog/data-analytics/how-to-find-outliers/
[*2]: https://www.investopedia.com/terms/e/empirical-rule.asp
[*3]: Adapted from: https://www.geeksforgeeks.org/z-score-for-outlier-detection-python/

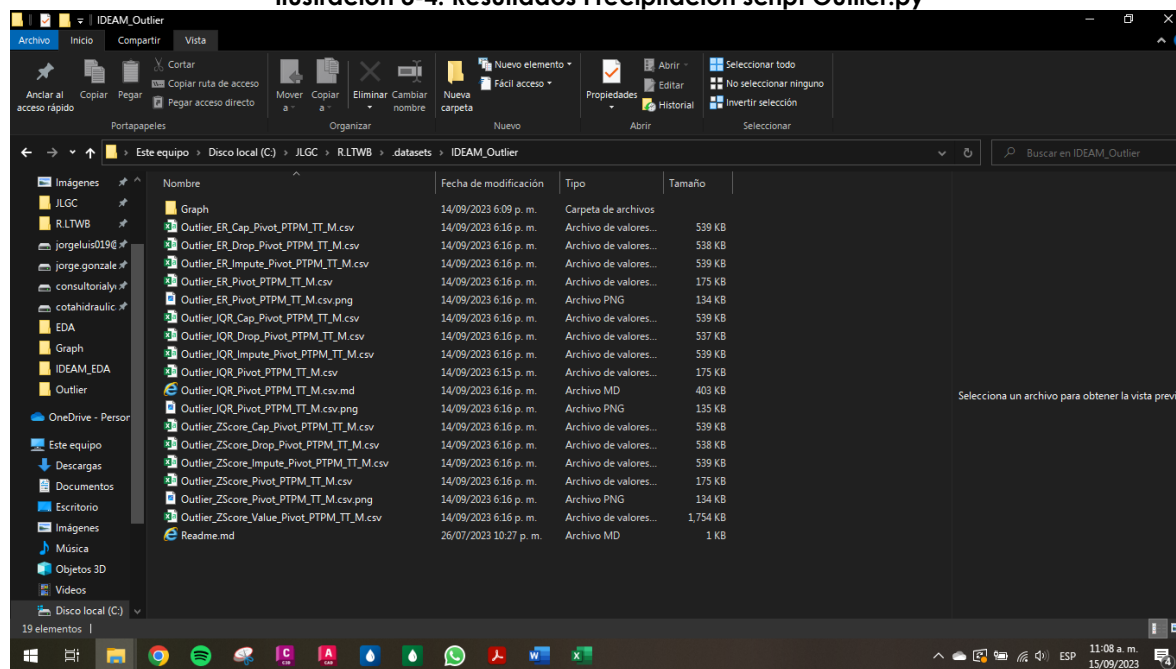
C:\JLGC\R.LTWB\datasets\IDEAM_Outlier>

```

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.

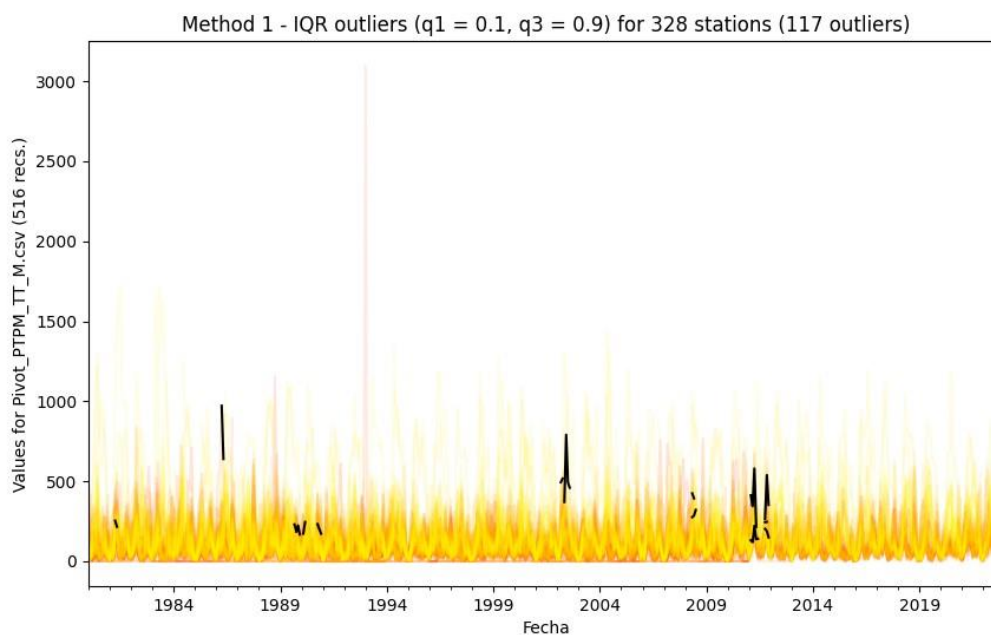
Ilustración 3-4. Resultados Precipitación script Outlier.py



Fuente: Elaboración Propia, 2023.

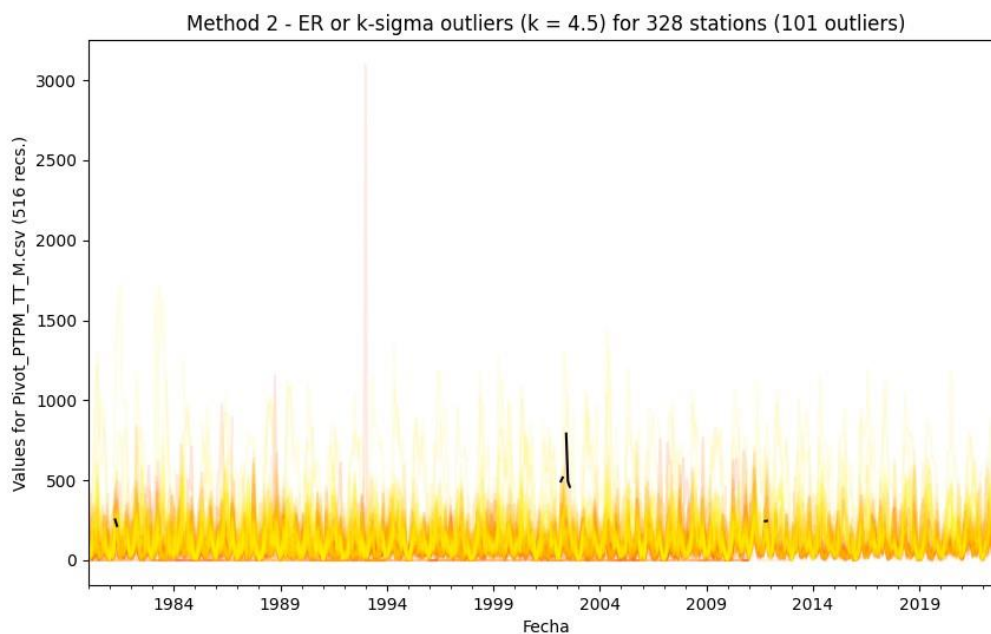
Las gráficas de resultados por cada método se presentan a continuación.

**Ilustración 3-5. Resultados Outlier Método 1 Precipitación**



**Fuente: Elaboración Propia, 2023.**

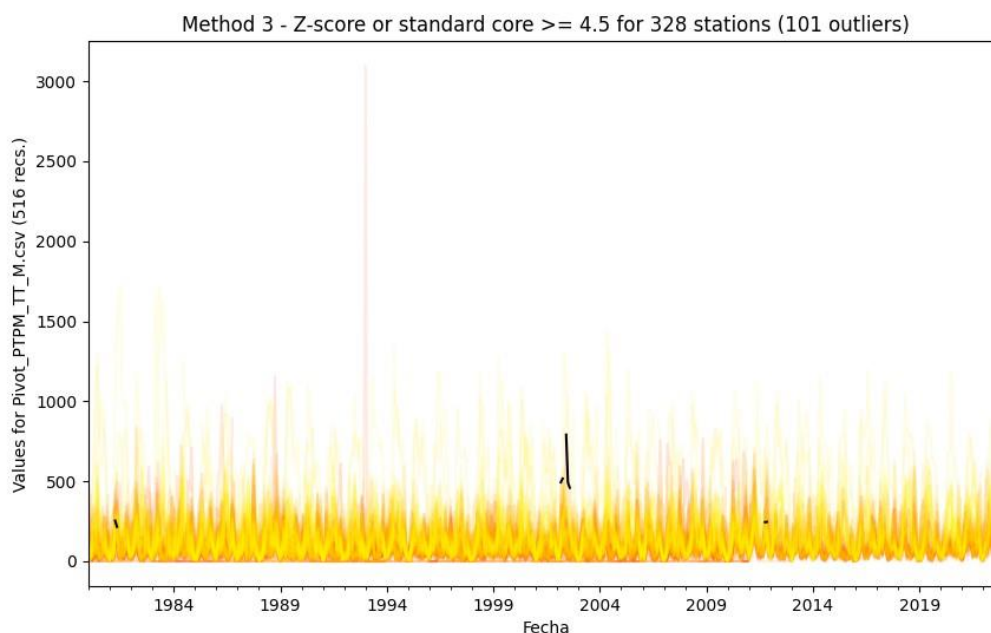
**Ilustración 3-6. Resultados Outlier Método 2 Precipitación**



**Fuente: Elaboración Propia, 2023.**



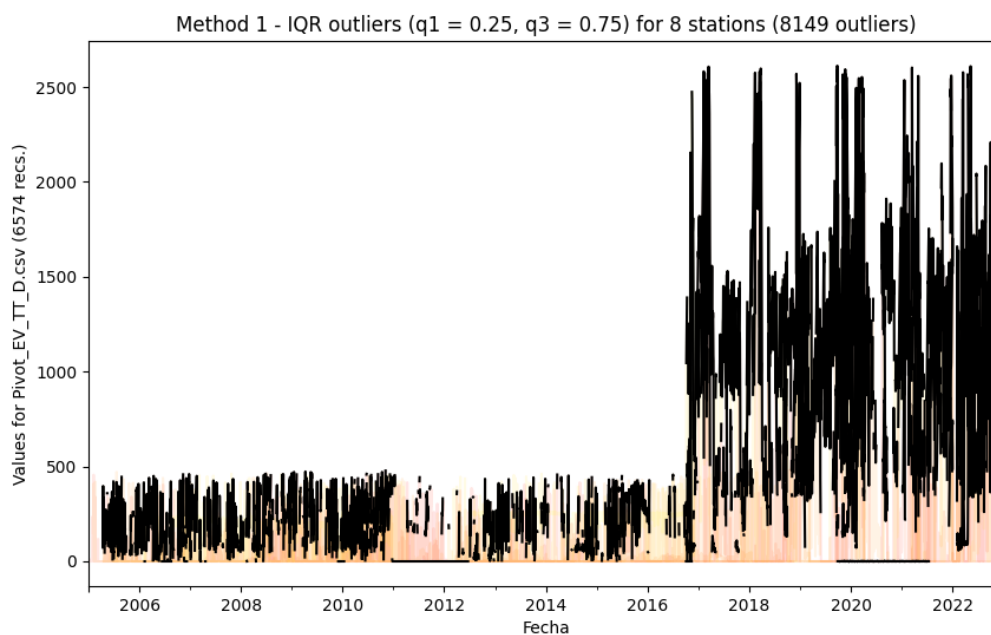
**Ilustración 3-7. Resultados Outlier Método 3 Precipitación**



**Fuente: Elaboración Propia, 2023.**

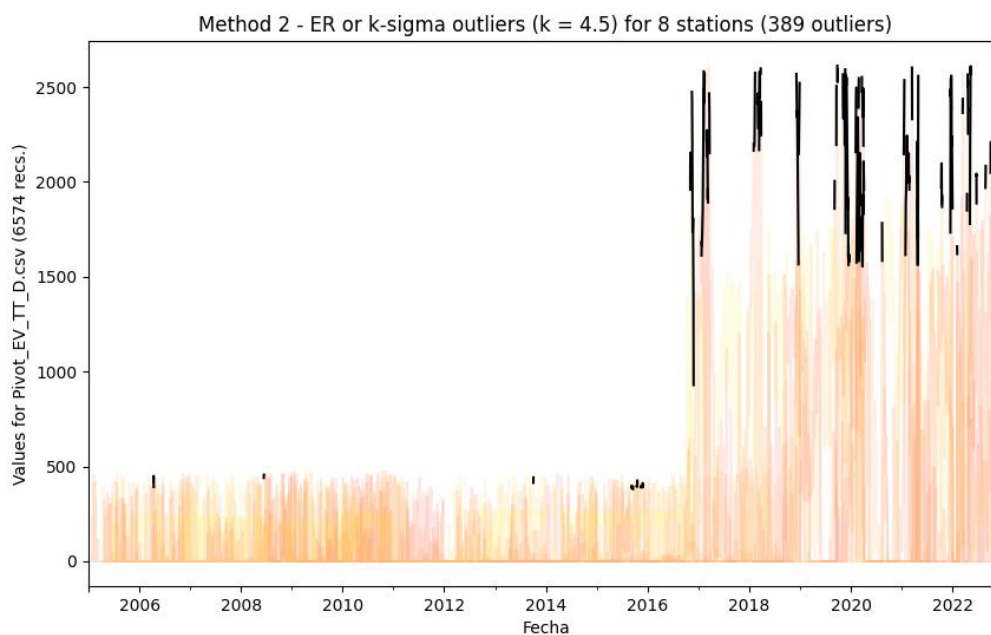
Se realizó el mismo ejercicio para los parámetros de caudal, evaporación, temperatura máxima y temperatura mínima.

**Ilustración 3-8. Resultados Outlier Método 1 Evaporación**



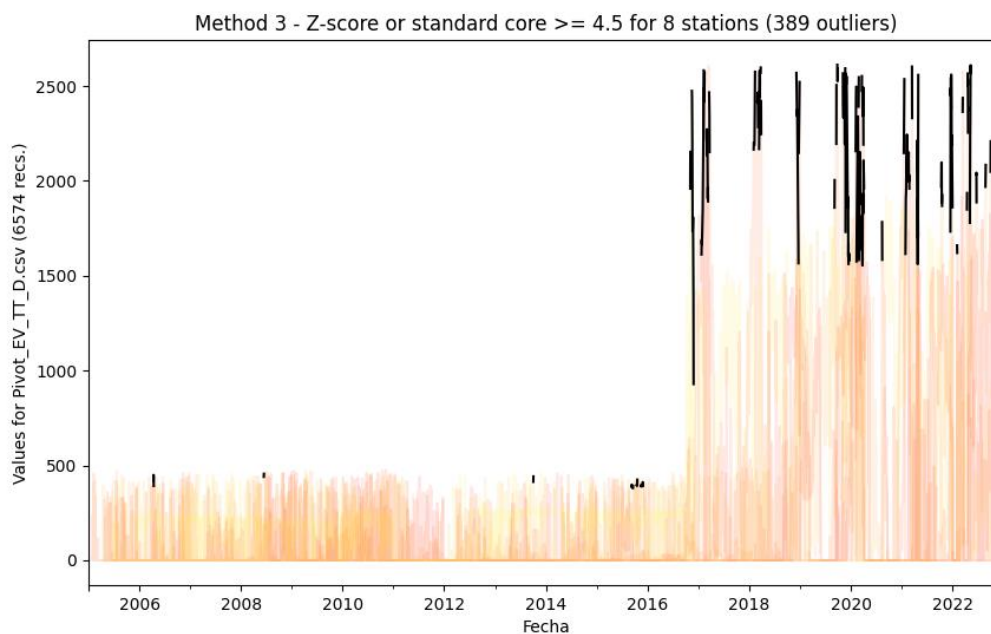
**Fuente: Elaboración Propia, 2023.**

**Ilustración 3-9. Resultados Outlier Método 2 Evaporación**



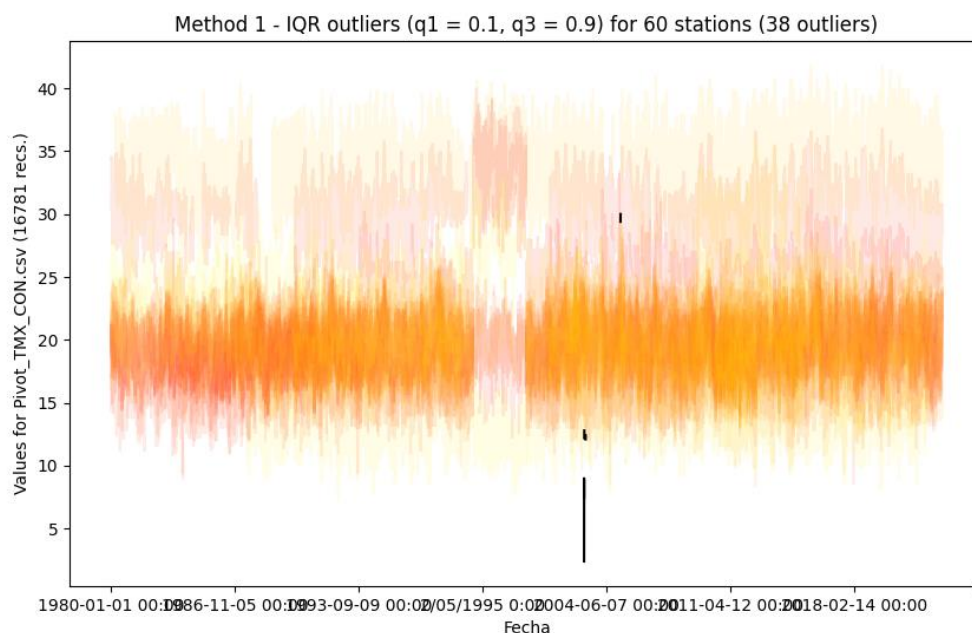
**Fuente: Elaboración Propia, 2023.**

**Ilustración 3-10. Resultados Outlier Método 3 Evaporación**



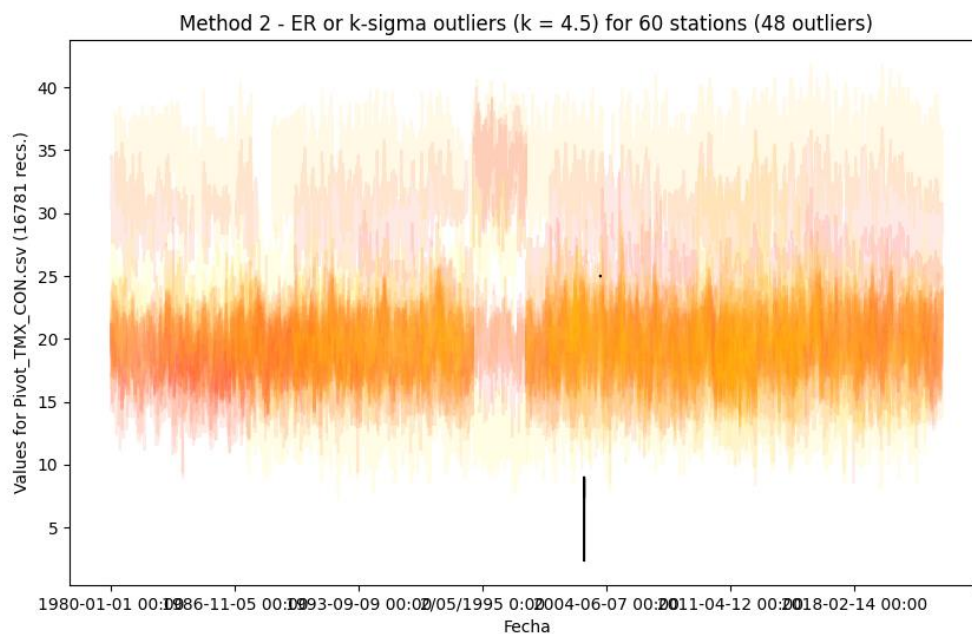
**Fuente: Elaboración Propia, 2023.**

**Ilustración 3-11. Resultados Outlier Método 1 Temperatura Máxima**



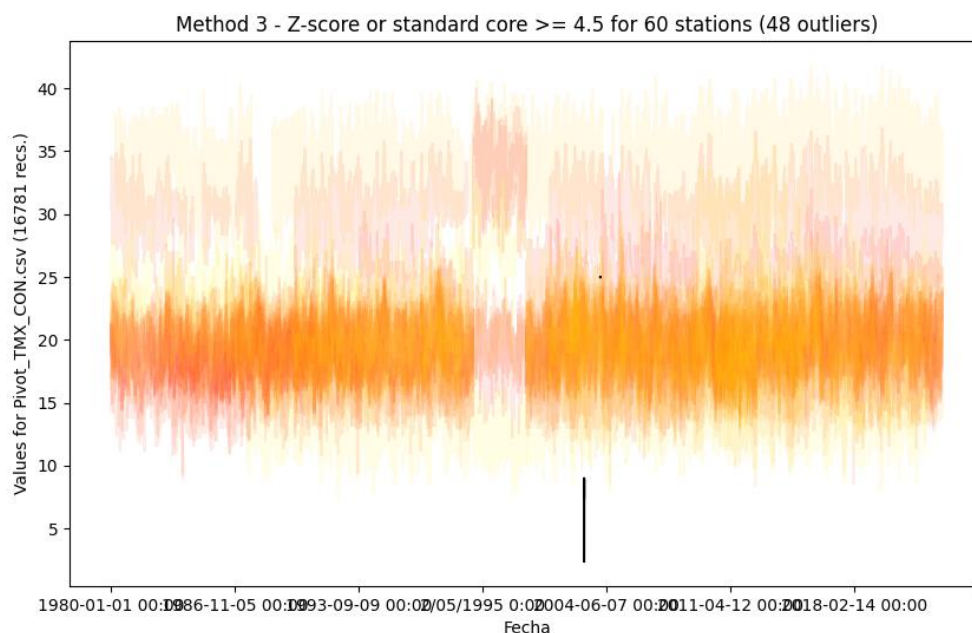
**Fuente: Elaboración Propia, 2023.**

**Ilustración 3-12. Resultados Outlier Método 2 Temperatura Máxima**



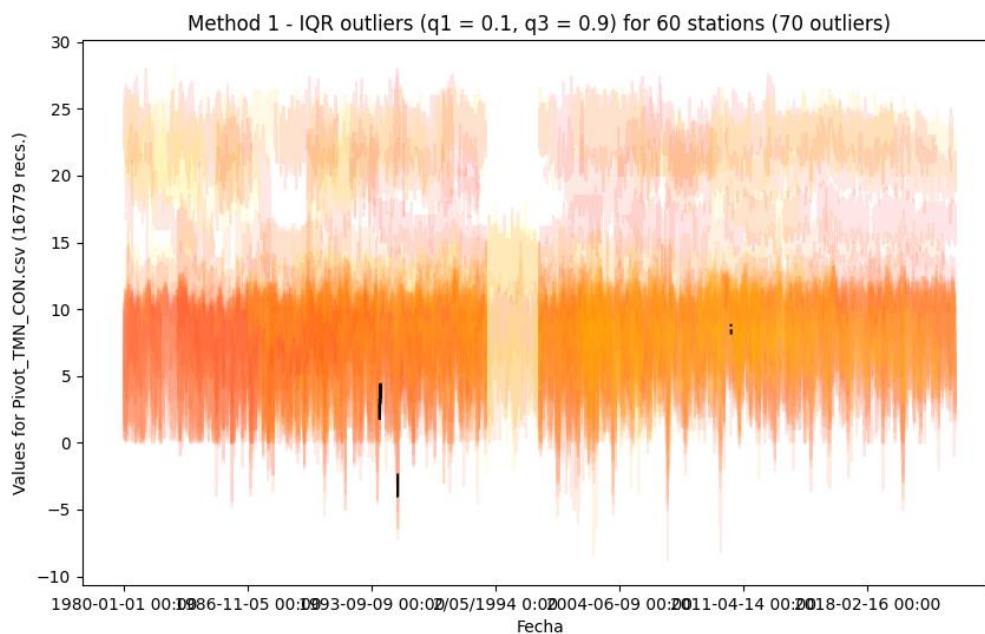
**Fuente: Elaboración Propia, 2023.**

**Ilustración 3-13. Resultados Outlier Método 3 Temperatura Máxima**



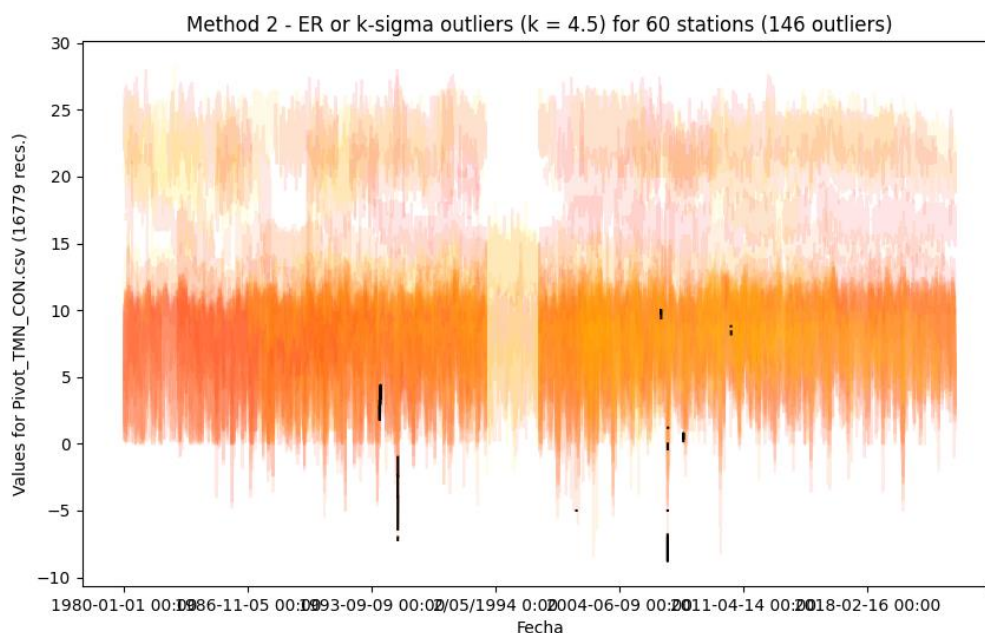
**Fuente: Elaboración Propia, 2023.**

**Ilustración 3-14. Resultados Outlier Método 1 Temperatura Mínima**



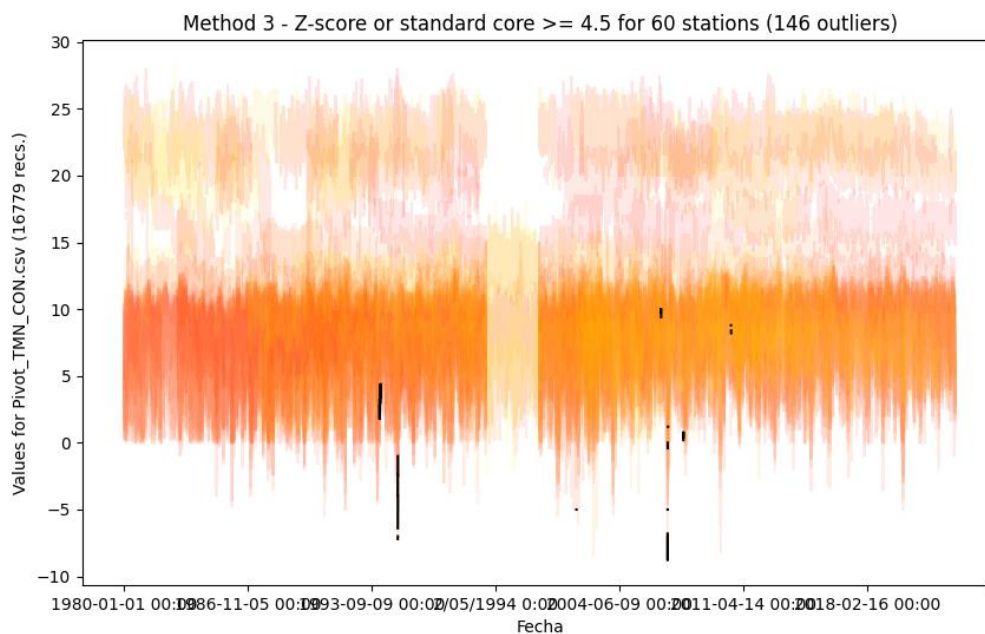
**Fuente: Elaboración Propia, 2023.**

**Ilustración 3-15. Resultados Outlier Método 2 Temperatura Mínima**



**Fuente: Elaboración Propia, 2023.**

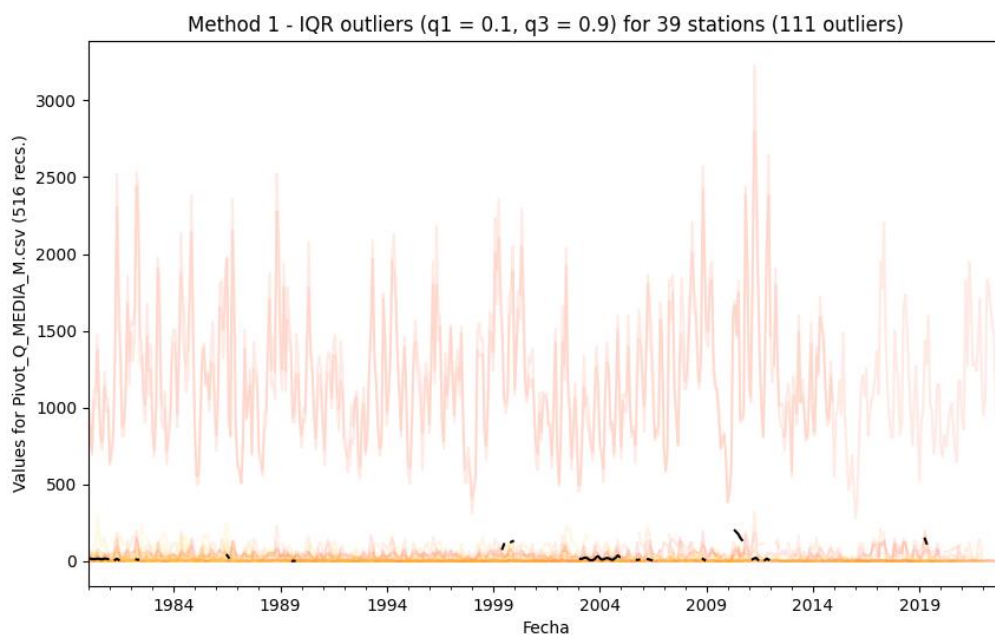
**Ilustración 3-16. Resultados Outlier Método 3 Temperatura Mínima**



**Fuente: Elaboración Propia, 2023.**

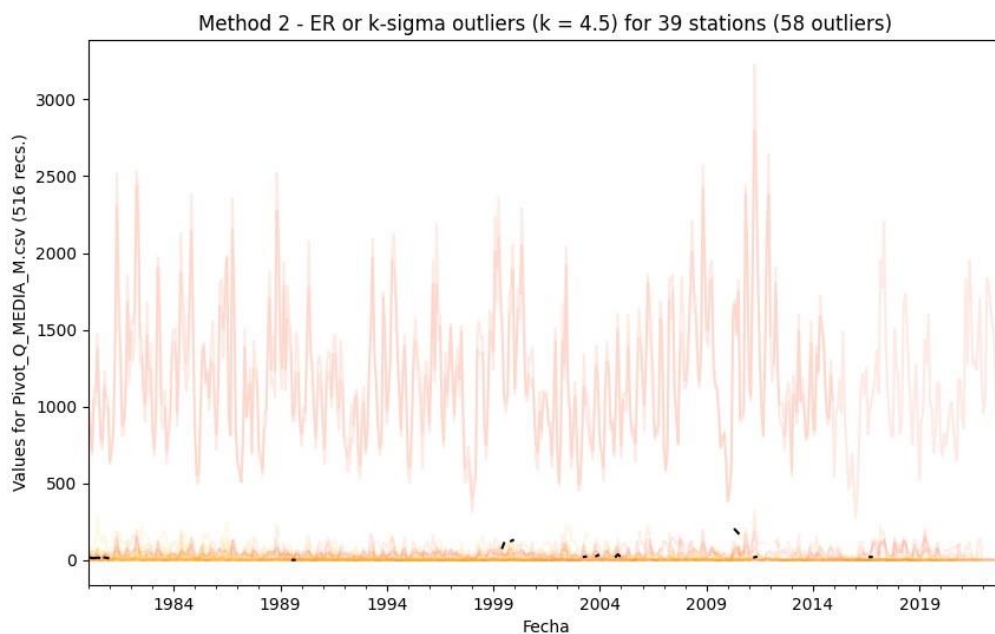


**Ilustración 3-17. Resultados Outlier Método 1 Caudal**



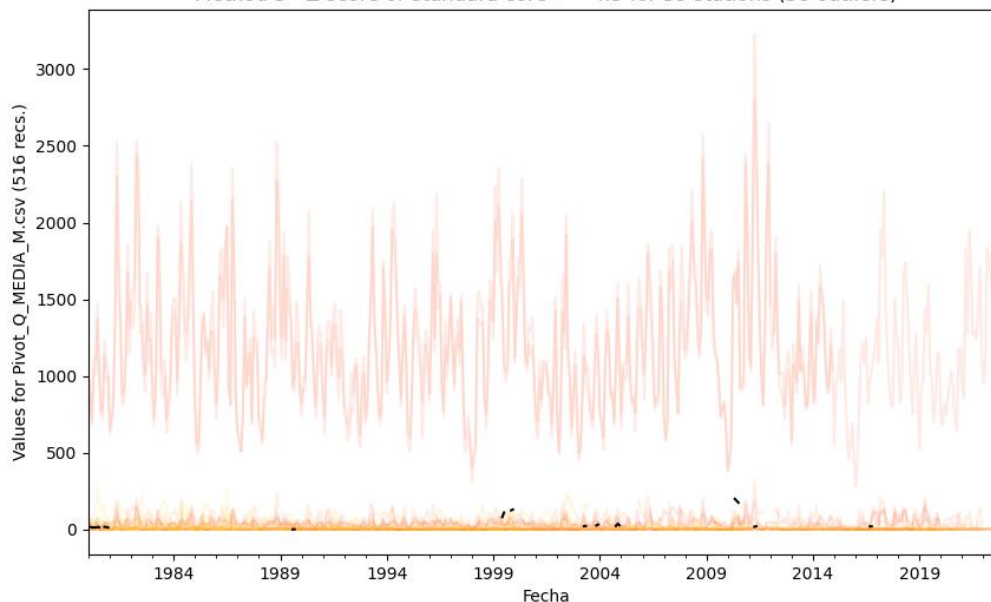
**Fuente: Elaboración Propia, 2023.**

**Ilustración 3-18. Resultados Outlier Método 2 Caudal**



**Fuente: Elaboración Propia, 2023.**

**Ilustración 3-19. Resultados Outlier Método 3 Caudal**  
Method 3 - Z-score or standard core  $\geq 4.5$  for 39 stations (58 outliers)



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 3-20. Registros Precipitación**

Estación	Serie Original		Método 1 - Valores Reemplazados		Método 1 - Valores Imputados		Método 2 - Valores Reemplazados		Método 2 - Valores Imputados		Método 3 - Valores Reemplazados		Método 3 - Valores Imputados		Análisis Desviación Estándar	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
21170020	100.962	77.354	100.962	77.354	100.962	77.354	100.962	77.354	100.962	77.354	100.962	77.354	100.962	77.354	0	1.5349E-14
21185040	100.744	77.636	100.640	77.132	99.951	75.548	100.640	77.132	99.951	75.548	100.640	77.132	99.951	75.548	0.383948285	0.93135995
21190030	96.842	67.474	96.670	66.594	95.967	64.962	96.670	66.594	95.967	64.962	96.670	66.594	95.967	64.962	0.403406806	1.03765939
21190060	131.326	90.194	131.326	90.194	131.326	90.194	131.326	90.194	131.326	90.194	131.326	90.194	131.326	90.194	0	0
21190110	73.958	33.612	73.958	33.612	73.958	33.612	73.958	33.612	73.958	33.612	73.958	33.612	73.958	33.612	0	7.6747E-15
21190170	98.029	81.359	98.029	81.359	98.029	81.359	98.029	81.359	98.029	81.359	98.029	81.359	98.029	81.359	0	0
21190210	122.542	108.166	121.463	100.298	120.496	97.901	121.463	100.298	120.496	97.901	121.463	100.298	120.496	97.901	0.763246075	3.63024992
21190300	174.428	153.124	174.428	153.124	174.428	153.124	174.428	153.124	174.428	153.124	174.428	153.124	174.428	153.124	3.0699E-14	0
21190310	142.807	89.402	142.807	89.402	142.807	89.402	142.807	89.402	142.807	89.402	142.807	89.402	142.807	89.402	0	1.5349E-14
21190360	217.380	111.237	217.380	111.237	217.380	111.237	217.380	111.237	217.380	111.237	217.380	111.237	217.380	111.237	6.13979E-14	1.5349E-14
21190430	70.306	87.225	70.306	87.225	70.306	87.225	70.306	87.225	70.306	87.225	70.306	87.225	70.306	87.225	1.53495E-14	1.5349E-14
21190440	74.457	103.899	74.457	103.899	74.457	103.899	74.457	103.899	74.457	103.899	74.457	103.899	74.457	103.899	0	0
21190450	56.991	71.786	56.991	71.786	56.991	71.786	56.991	71.786	56.991	71.786	56.991	71.786	56.991	71.786	1.53495E-14	0
21190460	97.794	58.301	97.602	57.330	96.725	55.269	97.602	57.330	96.725	55.269	97.602	57.330	96.725	55.269	0.499301293	1.27841618
21190470	76.261	48.122	76.261	48.122	76.261	48.122	76.261	48.122	76.261	48.122	76.261	48.122	76.261	48.122	0	7.6747E-15
21195030	87.462	58.519	87.462	58.519	87.462	58.519	87.462	58.519	87.462	58.519	87.462	58.519	87.462	58.519	1.53495E-14	0
21195030	170.093	89.911	170.093	89.911	170.093	89.911	170.093	89.911	170.093	89.911	170.093	89.911	170.093	89.911	1.53495E-14	1.5349E-14

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 3-21. Análisis Desviaciones Estándar Precipitación**

Estación	Serie Original		Método 1 - Valores Reemplazados		Método 1 - Valores Imputados		Método 2 - Valores Reemplazados		Método 2 - Valores Imputados		Método 3 - Valores Reemplazados		Método 3 - Valores Imputados		Análisis Desviación Estándar	
	mea	std	mea	std	mea	std	mea	std	mea	std	mea	std	mea	std	mean	std
21200710	66.205	72.622	66.021	71.747	63.526	65.719	66.021	71.747	63.526	65.719	66.021	71.747	63.526	65.719	1.359670392	3.35340906
21200800	67.413	68.151	65.513	53.570	61.738	40.986	65.503	53.513	62.672	44.429	65.503	53.513	62.672	44.429	2.073582947	9.12560141
21201090	130.214	125.314	129.132	119.133	122.365	101.531	128.918	118.158	124.406	106.678	128.918	118.158	124.406	106.678	3.085311106	8.6683583
21201130	39.284	214.980	40.559	137.867	22.219	29.832	29.378	77.817	24.793	38.554	29.378	77.817	24.793	38.554	7.224932512	67.1953031
21201380	65.632	54.125	65.388	52.933	63.514	48.349	65.388	52.933	63.514	48.349	65.388	52.933	63.514	48.349	1.037609778	2.64355761
21201620	57.139	54.803	56.803	52.828	51.929	39.459	56.621	52.012	53.697	44.380	56.621	52.012	53.697	44.380	2.065018026	5.75623153
21201720	59.824	80.735	59.222	76.954	52.157	57.356	58.861	75.341	54.824	64.665	58.861	75.341	54.824	64.665	2.964551981	8.50321309
21201750	80.825	89.550	82.167	95.199	70.973	65.766	80.825	89.550	80.825	89.550	80.825	89.550	80.825	89.550	3.840948931	9.57939939
21201760	22.476	35.008	22.081	32.765	19.317	25.040	21.990	32.367	19.687	26.003	21.990	32.367	19.687	26.003	1.389320085	4.09174364
21201790	12.148	40.022	12.967	42.414	5.021	17.674	11.660	37.491	9.012	30.285	11.660	37.491	9.012	30.285	2.749447744	8.41421565
21201840	21.000	74.893	26.155	87.481	5.092	17.338	19.317	64.450	15.600	53.572	19.317	64.450	15.600	53.572	6.518298886	22.0483115
21206070	101.584	82.502	101.584	82.502	101.584	82.502	101.464	81.940	99.401	77.067	101.464	81.940	99.401	77.067	1.043154671	2.55466547
21230080	88.953	93.328	88.545	91.317	85.853	84.801	88.545	91.317	85.853	84.801	88.545	91.317	85.853	84.801	1.50052915	3.81804371
23060040	113.111	109.778	112.574	107.135	108.832	97.964	112.574	107.135	108.832	97.964	112.574	107.135	108.832	97.964	2.080623224	5.33787922
24010380	22.084	43.361	21.835	41.878	18.786	33.859	21.702	41.296	20.178	37.465	21.702	41.296	20.178	37.465	1.229621934	3.34450836

**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 3-22. 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 3-23. Análisis Desviaciones Estándar Evaporación**

Estación	Serie Original		Método 1 - Valores Reemplazados		Método 1 - Valores Imputados		Método 2 - Valores Reemplazados		Método 2 - Valores Imputados		Método 3 - Valores Reemplazados		Método 3 - Valores Imputados		Análisis Desviación Estándar	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
21195190	221.429	446.829	515.215	936.935	54.059	93.390	221.429	446.829	221.429	446.829	221.429	446.829	221.429	446.829	136.650295	245.782653
21206930	136.260	319.190	234.782	513.216	54.090	82.278	130.540	287.186	110.002	228.724	130.540	287.186	110.002	228.724	54.12813619	129.328663
21206950	160.363	440.671	424.357	846.462	37.614	63.953	157.197	424.978	132.866	362.770	157.197	424.978	132.866	362.770	119.2894404	229.361789
21206980	83.629	326.543	267.271	582.111	17.400	31.120	73.000	258.984	44.807	155.705	73.000	258.984	44.807	155.705	82.9679032	174.144155
21206990	178.095	368.500	283.834	608.300	67.838	104.040	177.641	366.321	172.571	354.615	177.641	366.321	172.571	354.615	62.39651223	145.703762
21255160	24.061	79.268	75.474	151.808	4.769	9.593	23.773	77.882	20.409	69.657	23.773	77.882	20.409	69.657	22.20497404	41.338038
24015110	281.441	478.338	471.994	884.418	118.626	176.767	281.441	478.338	281.441	478.338	281.441	478.338	281.441	478.338	102.2327868	205.866352
35035130	186.647	153.257	95.344	282.092	210.057	42.630	181.714	104.224	178.441	92.532	181.714	104.224	178.441	92.532	36.05737822	76.6280183

**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 3-24. Registros Caudal**

Estación	Serie Original		Método 1 - Valores Reemplazados		Método 1 - Valores Imputados		Método 2 - Valores Reemplazados		Método 2 - Valores Imputados		Método 3 - Valores Reemplazados		Método 3 - Valores Imputados		Análisis Desviación Estándar	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
21197080	1.862	1.600	1.859	1.584	1.833	1.523	1.859	1.584	1.833	1.523	1.859	1.584	1.833	1.523	0.014374159	0.0349889
21197110	3.168	2.272	3.168	2.272	3.168	2.272	3.168	2.272	3.168	2.272	3.168	2.272	3.168	2.272	4.79671E-16	0
21197150	46.722	33.494	46.722	33.494	46.722	33.494	46.722	33.494	46.722	33.494	46.722	33.494	46.722	33.494	0.161708033	0.37074715
21197210	0.131	0.370	0.128	0.306	0.077	0.065	0.114	0.256	0.080	0.074	0.114	0.256	0.080	0.074	0.023933783	0.12678287
21207960	41.697	23.080	41.673	22.962	40.876	21.062	41.662	22.911	41.130	21.663	41.662	22.911	41.130	21.663	0.34624407	0.83013167
21209200	54.858	29.852	54.858	29.852	54.858	29.852	54.858	29.852	54.858	29.852	54.858	29.852	54.858	29.852	7.67474E-15	3.8374E-15
21209920	5.186	1.815	5.186	1.815	5.186	1.815	5.186	1.815	5.186	1.815	5.186	1.815	5.186	1.815	0	0
21227010	5.119	4.621	5.069	4.235	4.926	3.861	5.064	4.211	5.017	4.090	5.064	4.211	5.017	4.090	0.061302428	0.22971399
21237010	1148.510	410.094	1148.510	410.094	1148.510	410.094	1148.510	410.094	1148.510	410.094	1148.510	410.094	1148.510	410.094	0	0
21237020	1242.840	452.952	1242.840	452.952	1242.840	452.952	1242.840	452.952	1242.840	452.952	1242.840	452.952	1242.840	452.952	0	1.228E-13
21237040	3.102	3.970	3.065	3.731	2.985	3.531	3.065	3.731	2.985	3.531	3.065	3.731	2.985	3.531	0.049461652	0.16249361
21257090	23.812	22.181	23.761	21.910	23.151	20.457	23.740	21.817	23.435	21.102	23.740	21.817	23.435	21.102	0.246362504	0.61039067
23017020	1.382	4.218	1.515	4.330	0.571	0.406	1.262	3.447	0.961	2.445	1.262	3.447	0.961	2.445	0.320458924	1.35311256
23017030	43.673	20.966	43.651	20.857	42.350	17.608	43.632	20.770	42.981	19.217	43.632	20.770	42.981	19.217	0.513689007	1.274726
23017060	5.308	2.837	5.301	2.801	5.230	2.633	5.301	2.801	5.230	2.633	5.301	2.801	5.230	2.633	0.039001425	0.09556318
23017090	37.229	17.167	37.229	17.167	37.229	17.167	37.229	17.167	37.229	17.167	37.229	17.167	37.229	17.167	0	0
35067050	65.790	46.888	65.696	46.417	65.255	45.397	65.696	46.417	65.255	45.397	65.696	46.417	65.255	45.397	0.250694062	0.63037781

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 3-25. Análisis Desviaciones Estándar Caudal**

Estación	Serie Original		Método 1 - Valores Reemplazados		Método 1 - Valores Imputados		Método 2 - Valores Reemplazados		Método 2 - Valores Imputados		Método 3 - Valores Reemplazados		Método 3 - Valores Imputados		Análisis Desviación Estándar	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
35067050	24.029	30.574	23.754	29.013	19.954	17.302	23.584	28.246	21.304	21.797	23.584	28.246	21.304	21.797	1.612778238	4.95782209

Fuente: Elaboración Propia, 2023.

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

**Ilustración 3-26. Registros Temperatura Mínima**

Estación	Serie Original		Método 1 - Valores Reemplazados		Método 1 - Valores Imputados		Método 2 - Valores Reemplazados		Método 2 - Valores Imputados		Método 3 - Valores Reemplazados		Método 3 - Valores Imputados		Análisis Desviación Estándar	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
21185040	23.122	1.345	23.122	1.345	23.122	1.345	23.122	1.345	23.122	1.345	23.122	1.345	23.122	1.345	3.83737E-15	2.3984E-16
21195080	21.693	1.532	21.694	1.527	21.697	1.521	21.694	1.527	21.697	1.521	21.694	1.527	21.697	1.521	0.001438749	0.00395933
21195120	16.271	1.244	16.271	1.244	16.271	1.244	16.271	1.244	16.271	1.244	16.271	1.244	16.271	1.244	0	0
21195190	11.385	1.530	11.388	1.513	11.406	1.471	11.388	1.513	11.406	1.471	11.388	1.513	11.406	1.471	0.010147202	0.02528452
21205012	9.415	2.242	9.417	2.234	9.423	2.220	9.417	2.234	9.425	2.217	9.417	2.234	9.425	2.217	0.004161617	0.01037536
21205160	8.227	3.455	8.227	3.455	8.227	3.455	8.227	3.455	8.227	3.455	8.227	3.455	8.227	3.455	0	4.7967E-16
21205230	8.892	2.358	8.892	2.356	8.889	2.349	8.892	2.356	8.889	2.349	8.892	2.356	8.889	2.349	0.001572182	0.00375156
21205420	6.967	2.795	6.967	2.795	6.967	2.795	6.967	2.795	6.967	2.795	6.967	2.795	6.967	2.795	0	4.7967E-16
21205520	7.260	2.779	7.260	2.779	7.260	2.779	7.260	2.778	7.262	2.772	7.260	2.778	7.262	2.772	0.001210218	0.00291096
21205580	9.054	1.398	9.054	1.398	9.054	1.398	9.054	1.398	9.054	1.398	9.054	1.398	9.054	1.398	0	2.3984E-16
21205660	19.620	2.133	19.620	2.132	19.622	2.129	19.621	2.131	19.636	2.096	19.621	2.131	19.636	2.096	0.007397522	0.01708304
21205670	13.485	1.118	13.485	1.118	13.485	1.118	13.485	1.118	13.485	1.118	13.485	1.118	13.485	1.118	1.91868E-15	0
21205700	7.731	2.257	7.731	2.257	7.731	2.257	7.731	2.257	7.731	2.257	7.731	2.257	7.731	2.257	9.59342E-16	0
21205710	8.901	1.941	8.901	1.941	8.901	1.941	8.901	1.940	8.903	1.935	8.901	1.940	8.903	1.935	0.001123697	0.00259107
21205720	7.141	1.765	7.141	1.765	7.141	1.765	7.141	1.765	7.141	1.765	7.141	1.765	7.141	1.765	9.59342E-16	0
21205740	7.998	2.378	7.998	2.378	7.998	2.378	7.998	2.378	8.001	2.372	7.998	2.378	8.001	2.372	0.001178599	0.00273505
21205770	6.477	2.693	6.477	2.693	6.477	2.693	6.477	2.693	6.477	2.693	6.477	2.693	6.477	2.693	0	4.7967E-16

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 3-27. Registros Temperatura Máxima**

Estación	Serie Original		Método 1 - Valores Reemplazados		Método 1 - Valores Imputados		Método 2 - Valores Reemplazados		Método 2 - Valores Imputados		Método 3 - Valores Reemplazados		Método 3 - Valores Imputados		Análisis Desviación Estándar	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
21185040	33.782	2.177	33.782	2.177	33.782	2.177	33.782	2.177	33.782	2.177	33.782	2.177	33.782	2.177	7.67474E-15	0
21195080	33.645	1.999	33.645	1.999	33.645	1.999	33.645	1.999	33.645	1.999	33.645	1.999	33.645	1.999	0	2.3984E-16
21195120	24.885	1.710	24.885	1.710	24.884	1.708	24.885	1.710	24.884	1.708	24.885	1.710	24.884	1.708	0.000495696	0.00115815
21195190	19.626	1.567	19.626	1.566	19.626	1.564	19.626	1.566	19.626	1.564	19.626	1.566	19.626	1.564	0.000335942	0.0008656
21205012	19.596	1.560	19.596	1.560	19.596	1.560	19.596	1.560	19.596	1.560	19.596	1.560	19.596	1.560	0	0
21205160	19.878	2.320	19.878	2.320	19.878	2.320	19.878	2.320	19.878	2.320	19.878	2.320	19.878	2.320	0	0
21205230	19.973	1.601	19.973	1.593	19.973	1.585	19.973	1.585	19.973	1.585	19.973	1.585	19.973	1.585	0.000113389	0.0063909
21205420	19.625	1.472	19.625	1.470	19.625	1.469	19.625	1.470	19.625	1.469	19.625	1.470	19.625	1.469	0.000275162	0.00100372
21205520	19.517	1.419	19.517	1.419	19.517	1.419	19.517	1.419	19.517	1.419	19.517	1.419	19.517	1.419	3.83737E-15	2.3984E-16
21205580	18.629	1.798	18.629	1.798	18.629	1.798	18.629	1.798	18.629	1.798	18.629	1.798	18.629	1.798	0	2.3984E-16
21205660	30.241	2.017	30.241	2.017	30.241	2.017	30.241	2.017	30.241	2.017	30.241	2.017	30.241	2.017	0	0
21205670	21.012	1.511	21.012	1.510	21.010	1.505	21.012	1.510	21.010	1.505	21.012	1.510	21.010	1.505	0.001290626	0.00305754
21205700	18.362	1.885	18.362	1.885	18.362	1.885	18.362	1.885	18.363	1.883	18.362	1.885	18.363	1.883	0.000341565	0.00084638
21205710	20.346	1.512	20.347	1.511	20.347	1.510	20.347	1.511	20.347	1.510	20.347	1.511	20.347	1.510	0.000354562	0.00092152
21205720	16.733	1.730	16.733	1.730	16.733	1.730	16.733	1.730	16.733	1.728	16.733	1.730	16.733	1.728	0.00029277	0.00071644
21205740	17.140	1.951	17.140	1.951	17.140	1.951	17.140	1.951	17.140	1.951	17.140	1.951	17.140	1.951	0	0
21205770	19.003	1.344	19.003	1.344	19.003	1.344	19.003	1.344	19.003	1.344	19.003	1.344	19.003	1.344	3.83737E-15	0

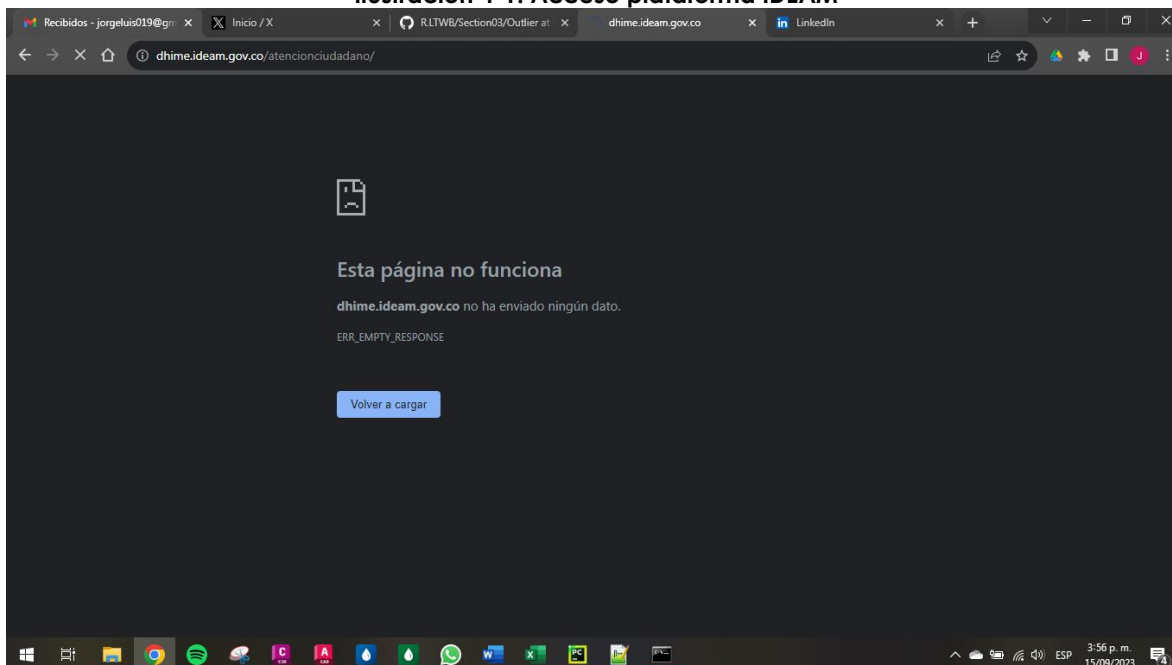
Fuente: Elaboración Propia, 2023.

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

## 4. 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 4-1. Acceso plataforma IDEAM**



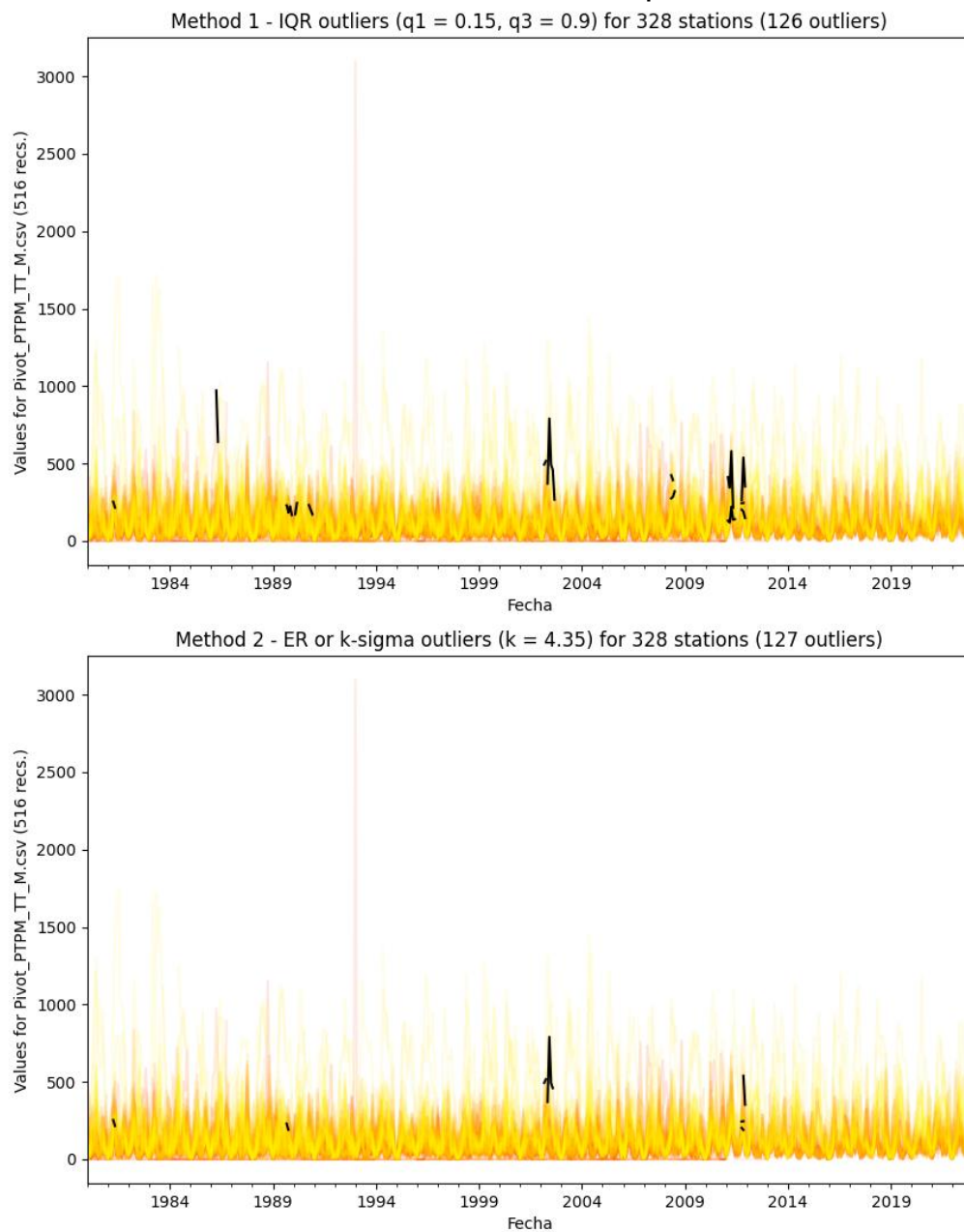
Fuente: IDEAM, 15/09/2023.

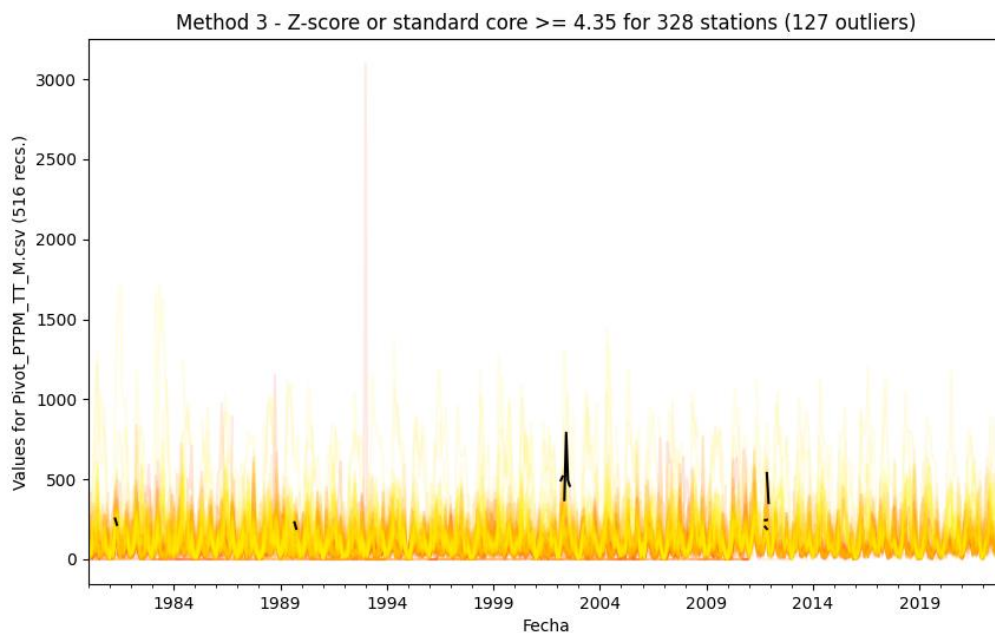
## 5. ACTIVIDAD 3: SENSIBILIDAD PARÁMETROS ESTADÍSTICOS

Se realizó un ejercicio de ensayo y error para determinar el  $q_1$ ,  $q_3$ , K-sigma y Z-score para generar un número similar de outliers.

Para la precipitación se tiene que  $q_1=0.15$ ,  $q_3=0.90$ , K-sigma=4.35 y Z-score=4.35.

**Ilustración 5-1. Resultados Precipitación**

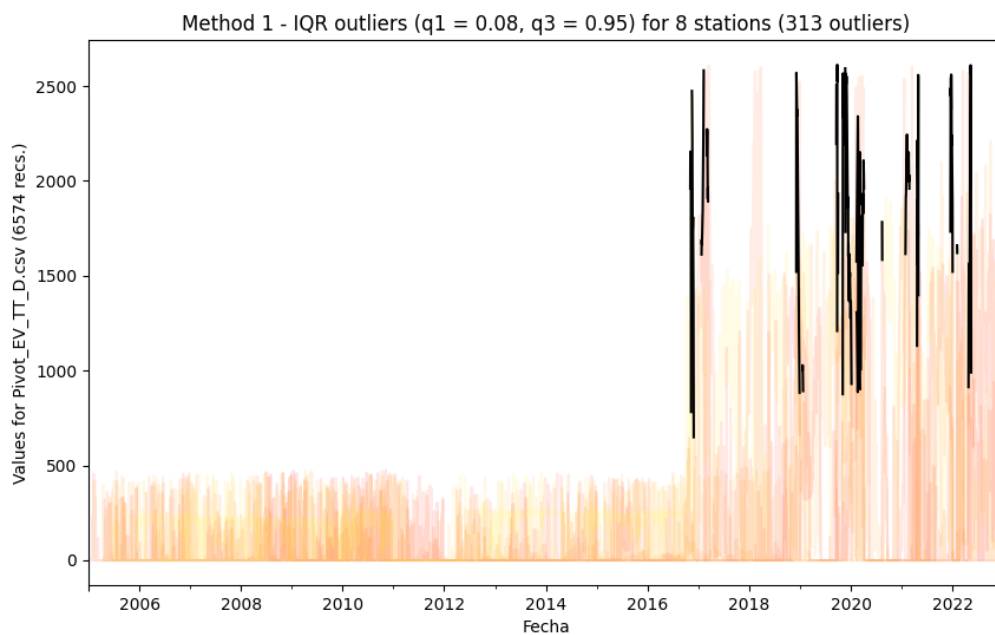




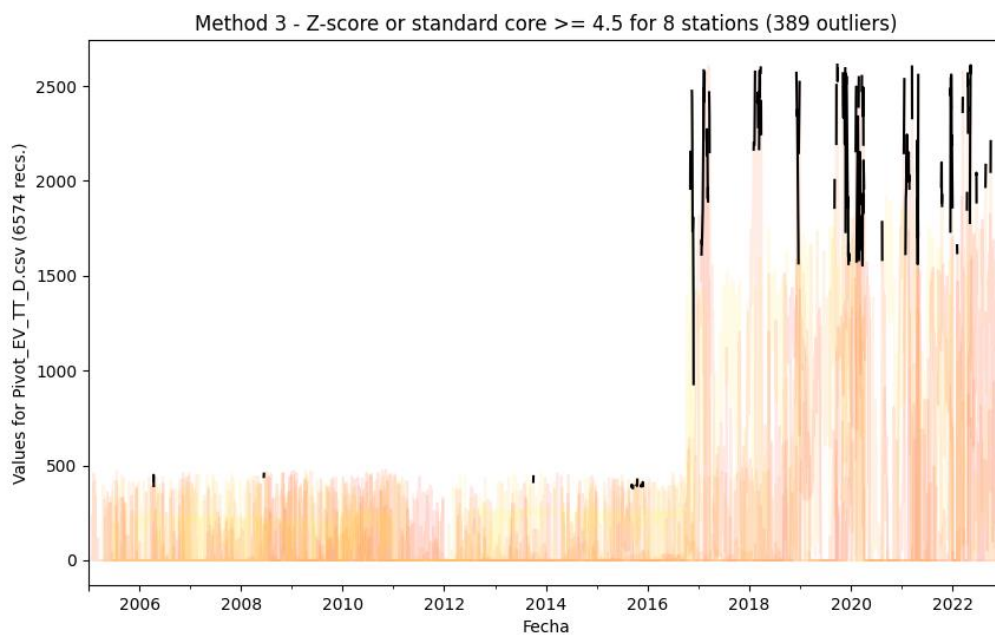
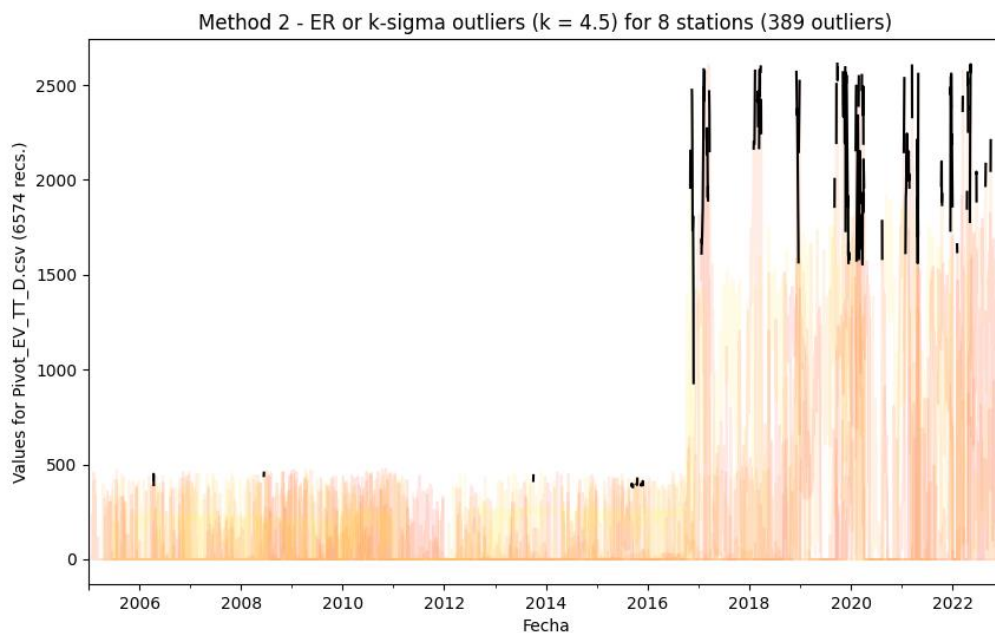
**Fuente: Elaboración Propia.**

Para evaporación se tiene que  $q1=0.08$ ,  $q3=0.95$ ,  $K\text{-sigma}=4.5$  y  $Z\text{-score}=4.5$ .

**Ilustración 5-2. Resultados Evaporación**



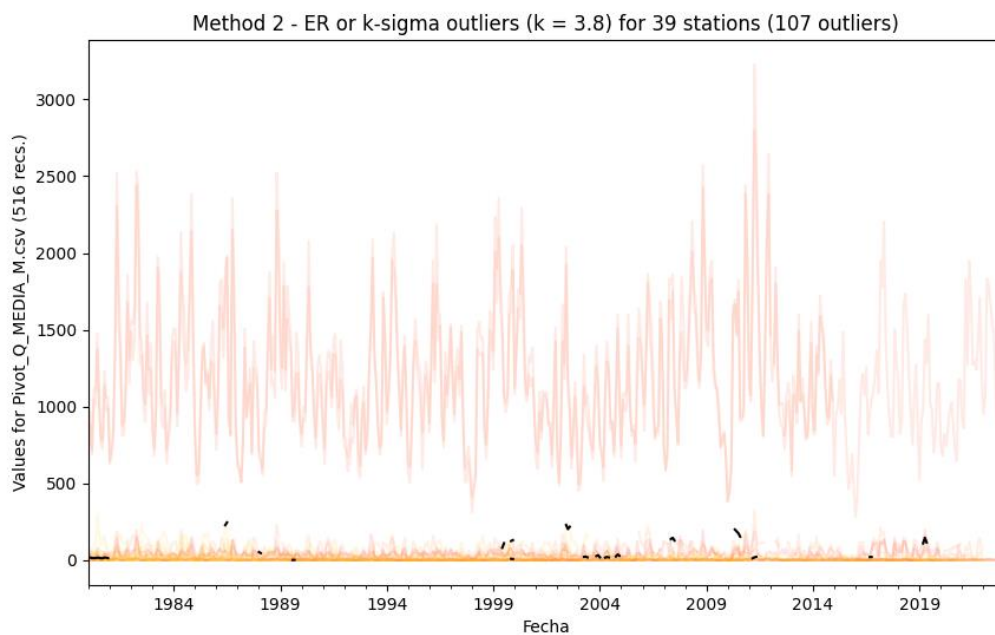
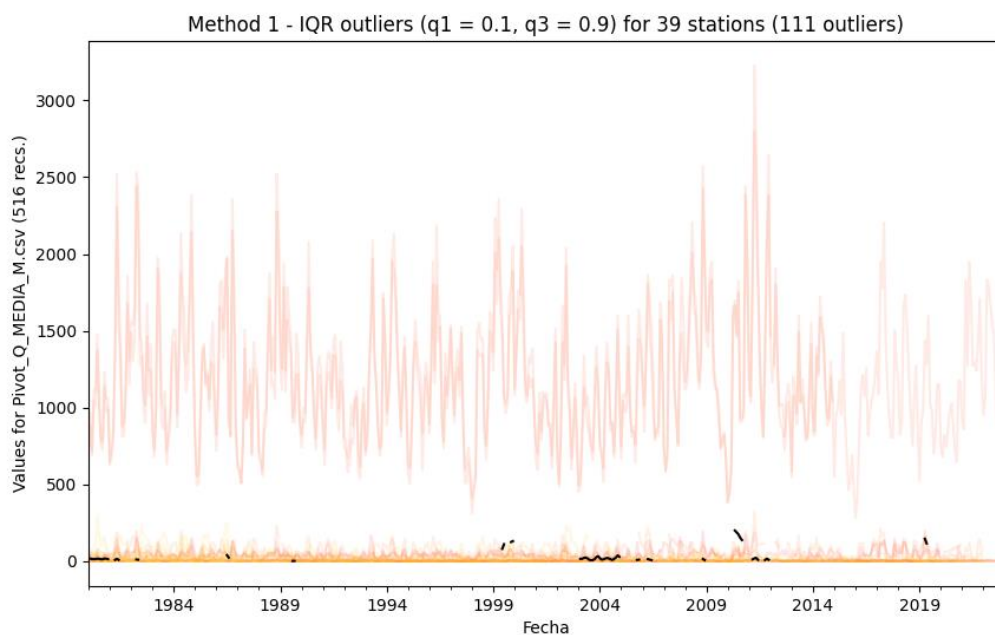




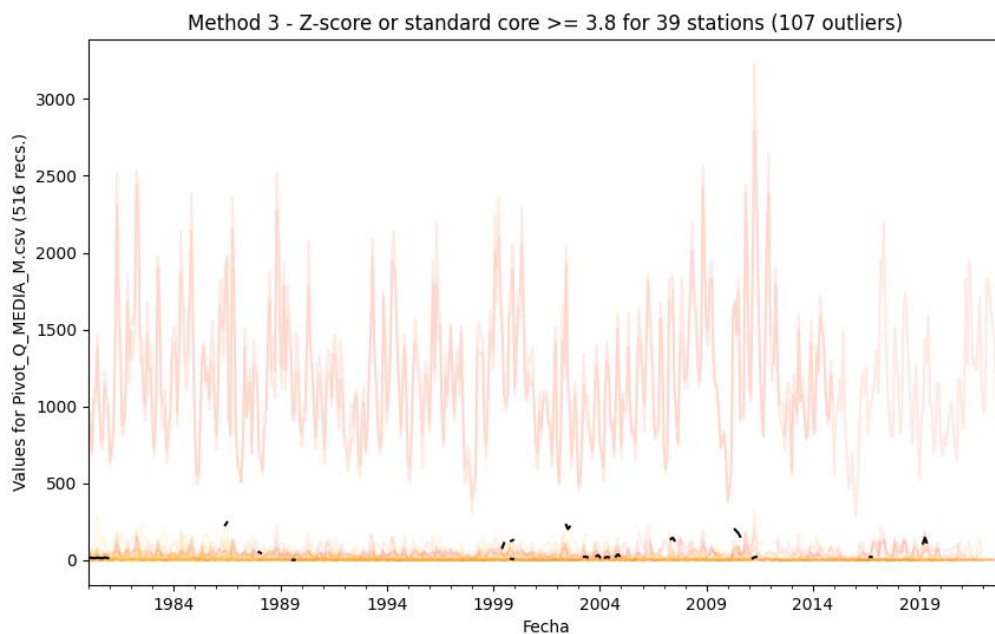
**Fuente: Elaboración Propia.**

Para caudal se tiene que  $q1=0.1$ ,  $q3=0.9$ ,  $K\text{-sigma}=3.8$  y  $Z\text{-score}=3.8$ .

**Ilustración 5-3. Resultados Caudal**



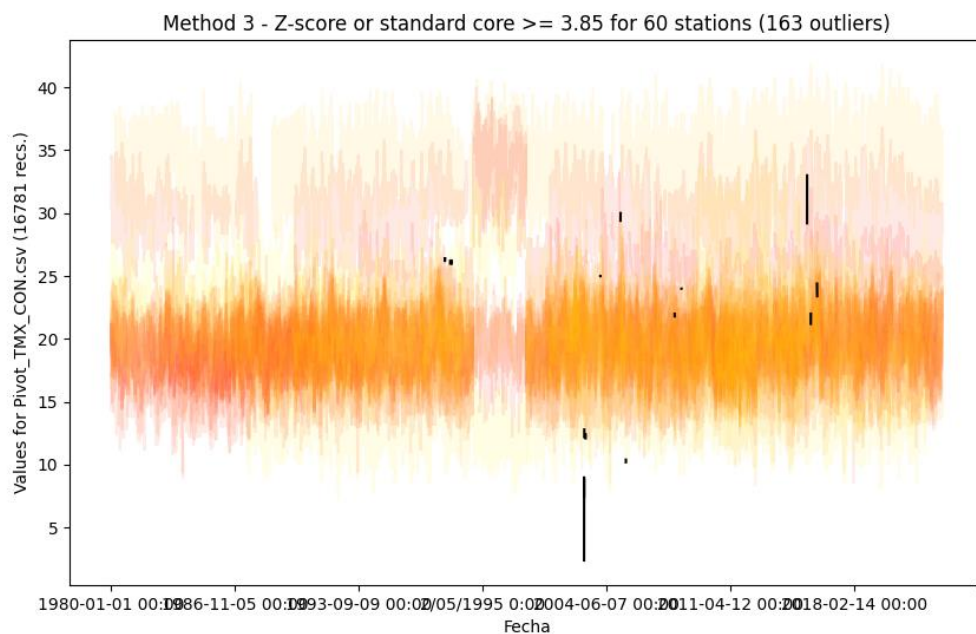


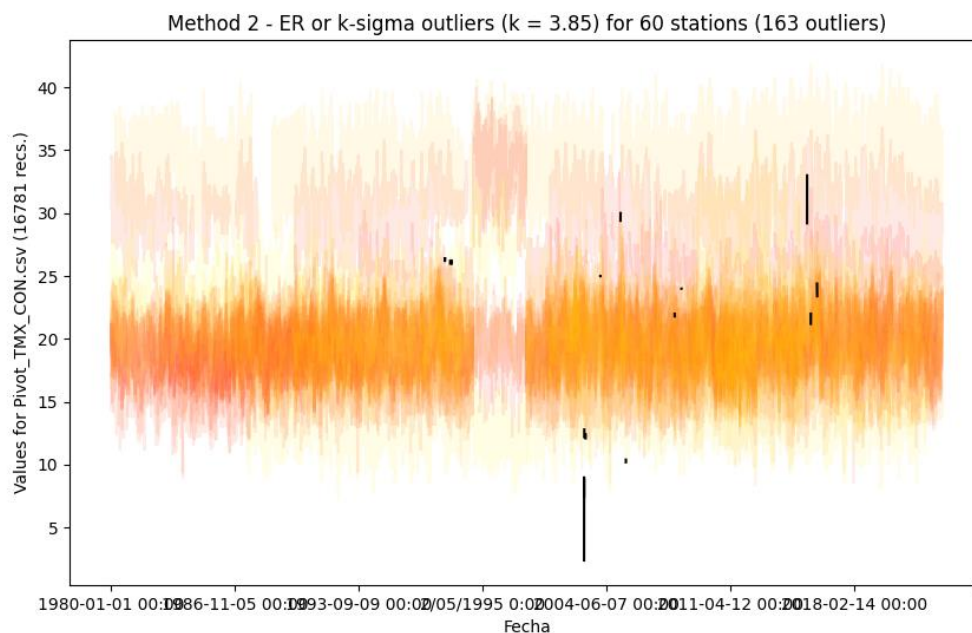
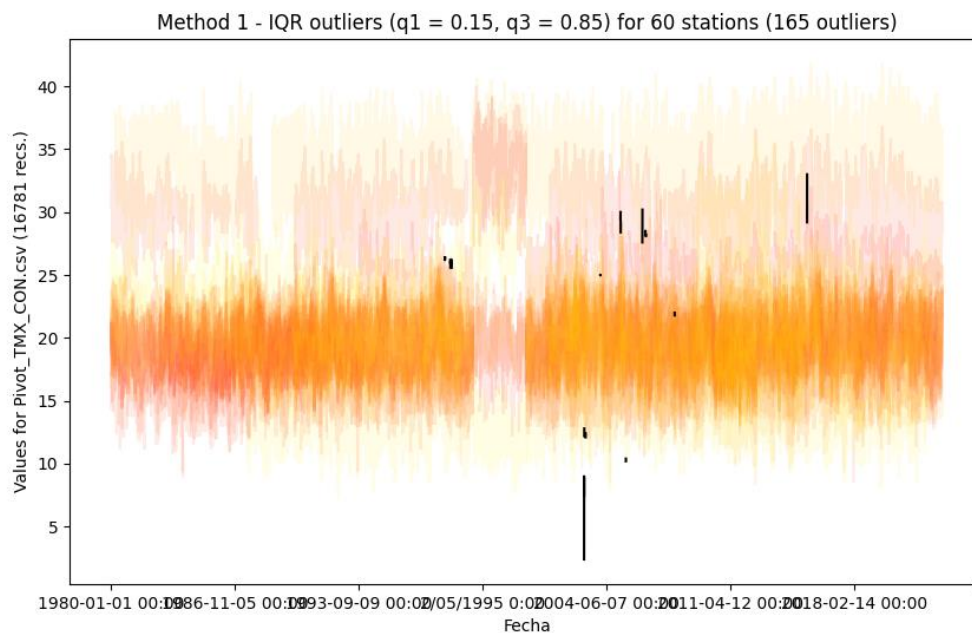


Fuente: Elaboración Propia.

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

**Ilustración 5-4. Resultados Temperatura Máxima**

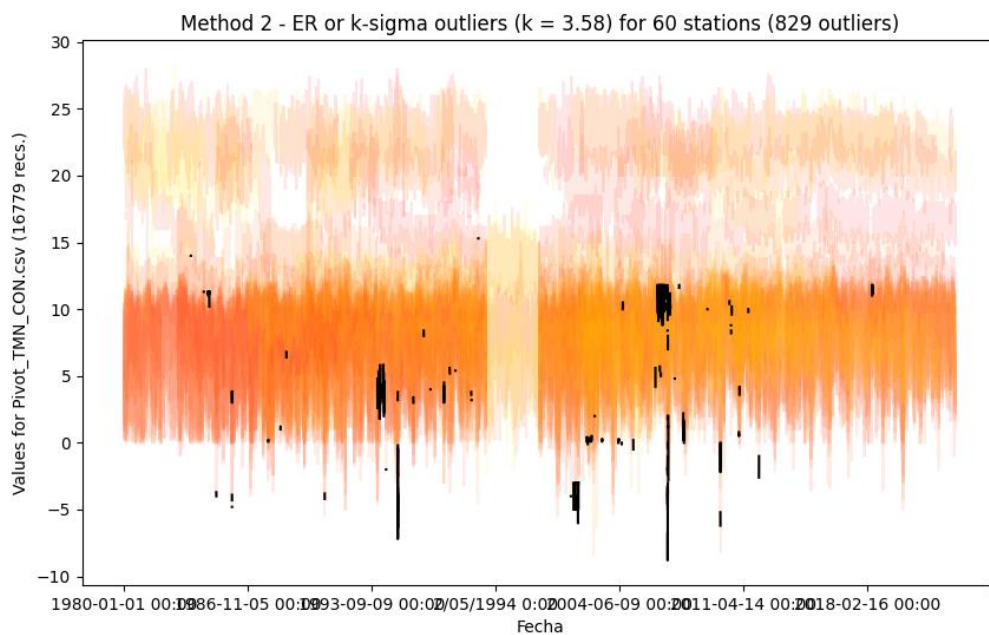
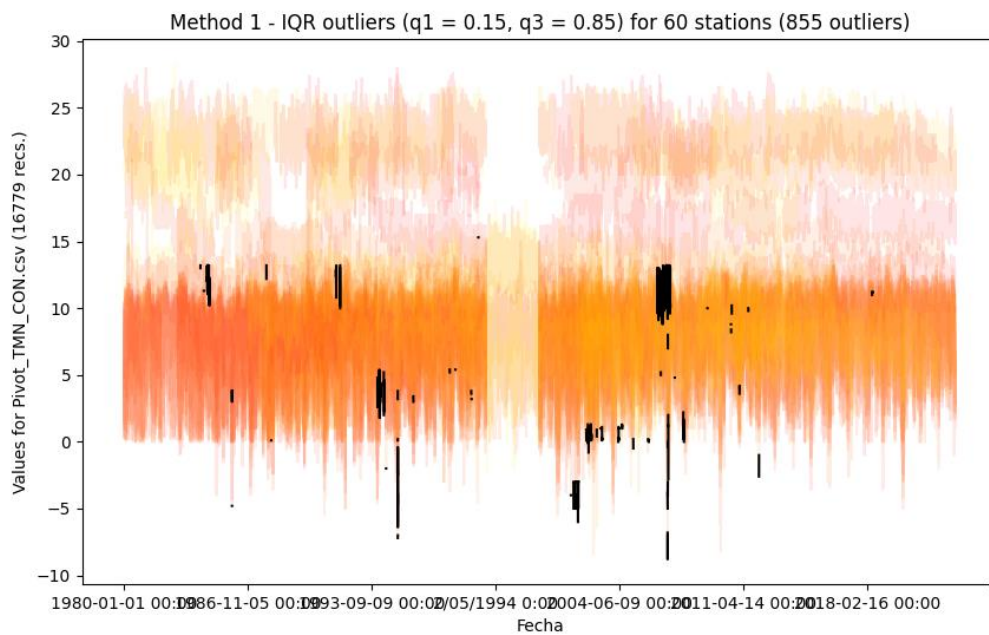


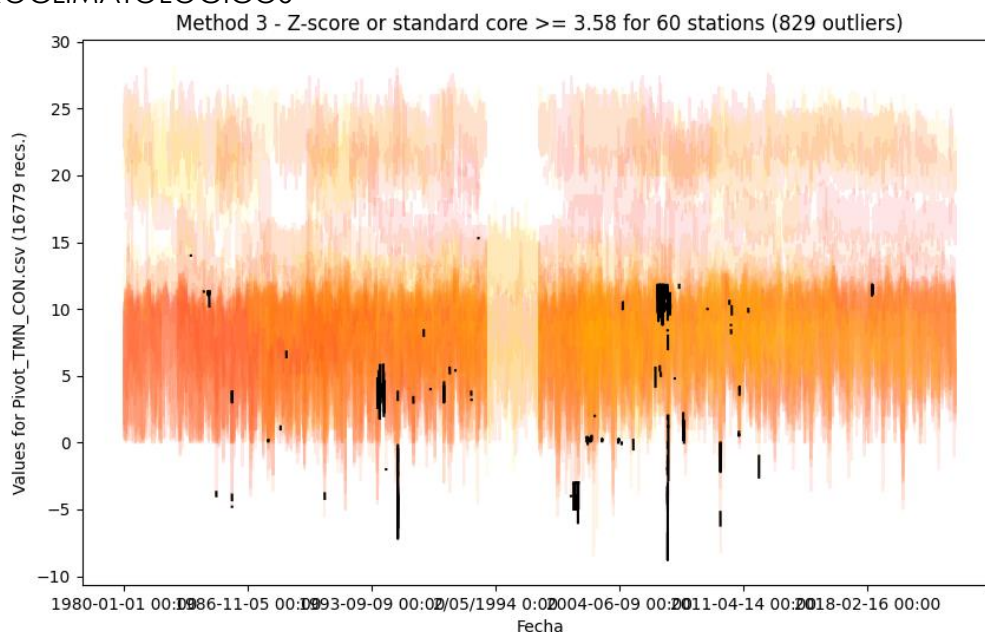


**Fuente: Elaboración Propia.**

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

**Ilustración 5-5. Resultados Temperatura Mínima**





Fuente: Elaboración Propia.

## 6. 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

## 7. 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 pudo 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.

## **8. 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>.