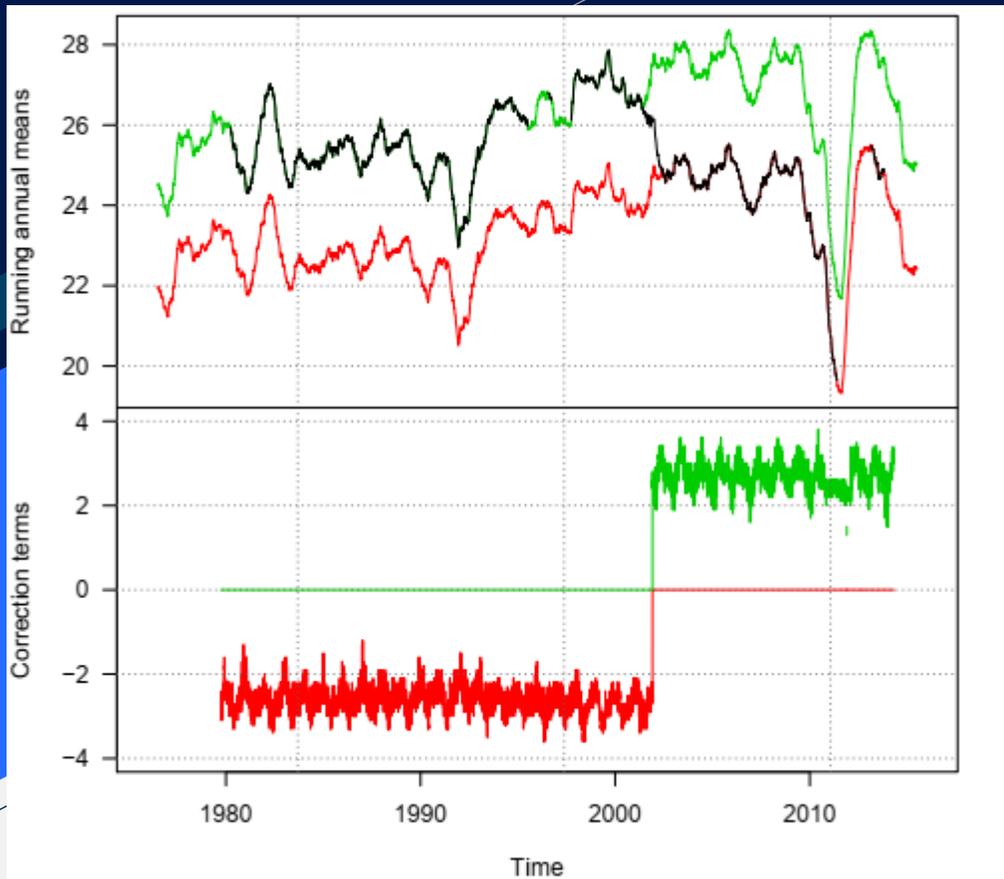




Webinar 19



La homogeneización de series climatológicas usando la herramienta CLIMATOL.

M.C. Óscar Pita-Díaz

Colaborador en la Subcoordinación de Hidrometeorología

Instituto Mexicano de Tecnología del Agua

18 de junio de 2020

Contenido

1. Generalidades

1.1 Introducción

1.2 Antecedentes de herramientas para la homogeneización de series de observaciones meteorológicas

1.3 Paquete de R *Climatol*

2. Reconstrucción de series climáticas con Climatol

2.1 Análisis exploratorio de los datos

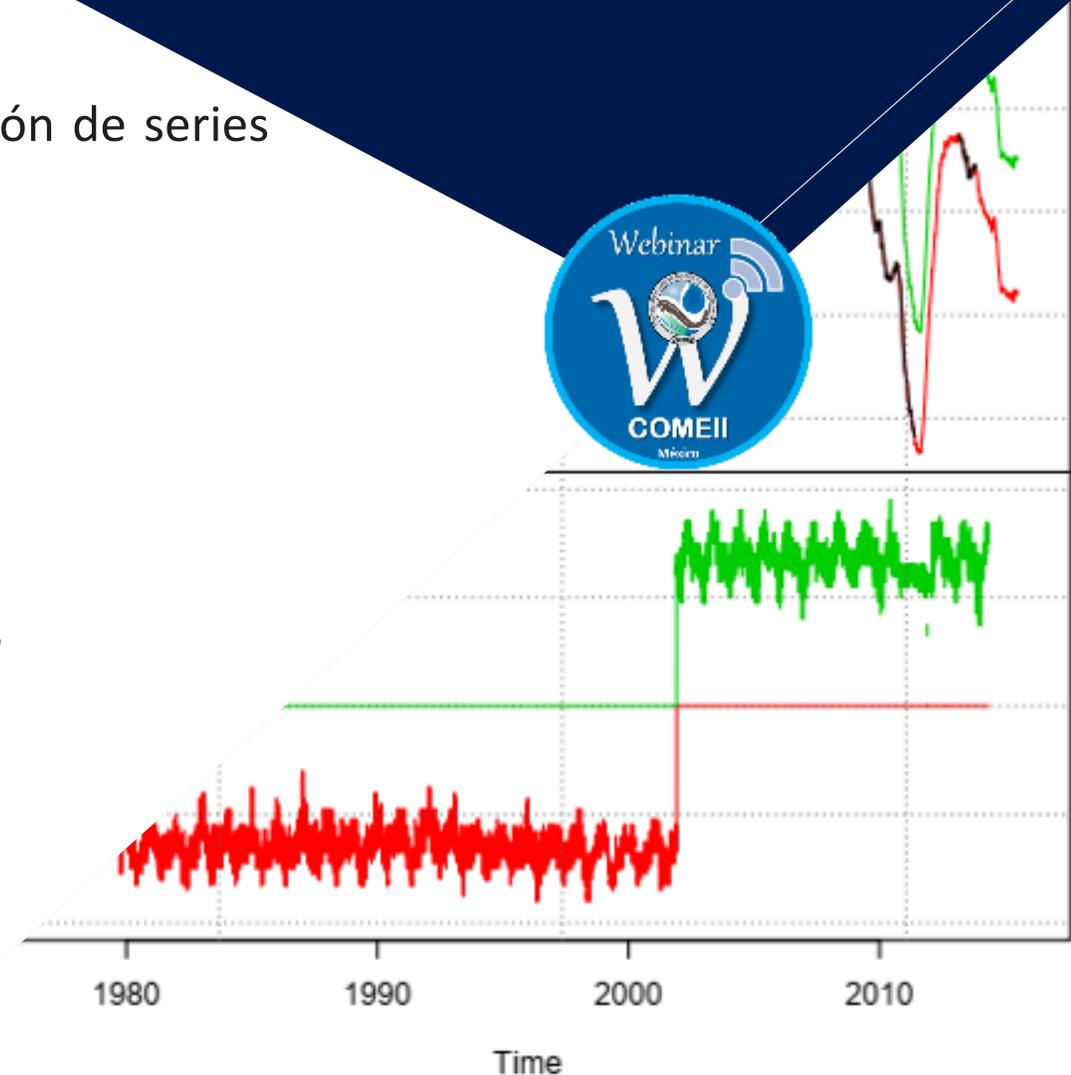
2.2 Homogeneización de series mensuales

2.3 Homogeneización de series diarias

2.4 Obtención de productos con los datos homogeneizados

2.5 Obtención de mallas homogeneizadas

3. Conclusiones



1.1 Introducción



- La palabra homogeneización se deriva de la palabra homogéneo que significa de la misma naturaleza. En climatología el termino se refiere al **proceso de eliminación de perturbaciones que no son meramente climáticas en los registros de los datos crudos.**

Guijarro, 2016

- Consiste en corregir los cambios del comportamiento de la serie debido a factores externos no climatológicos.

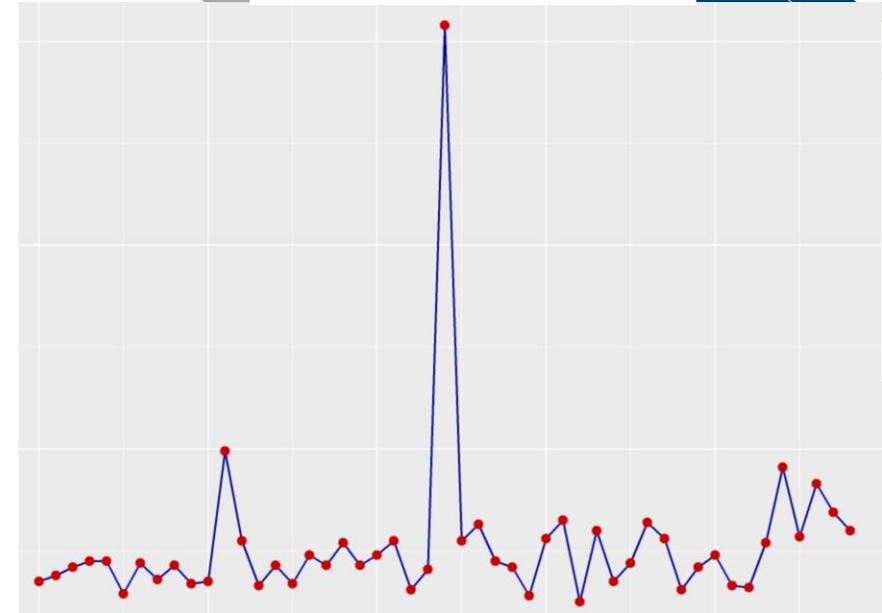
- Principales **problemas en las series de datos:**

- Ausencia de datos
- Valores sospechosos
- Comportamiento o tendencia sospechosa

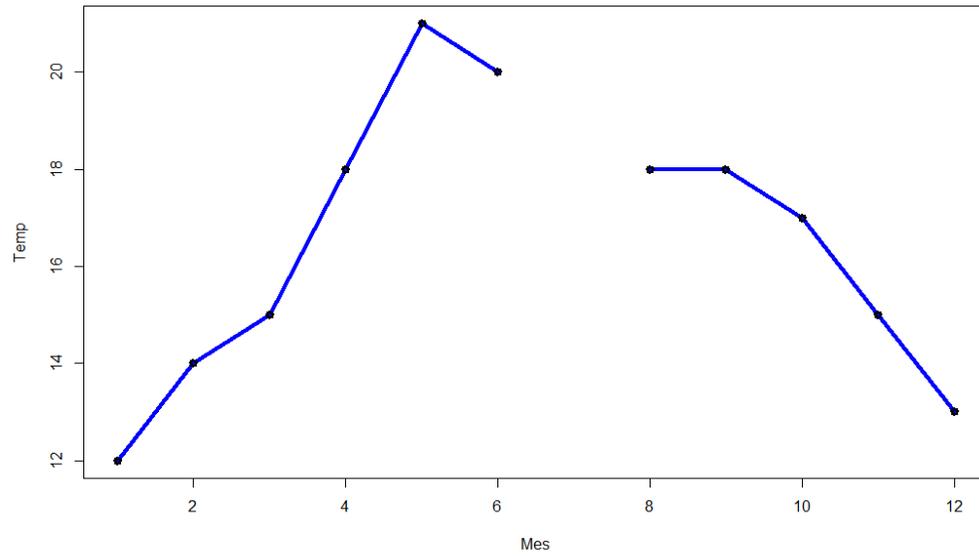
- Principales causas que originan **valores sospechosos:**

- Error en la medición
- Error de captura
- Problemas con la instrumentación (Recalibración, cambio de instrumento, cambio de ubicación de la estación)

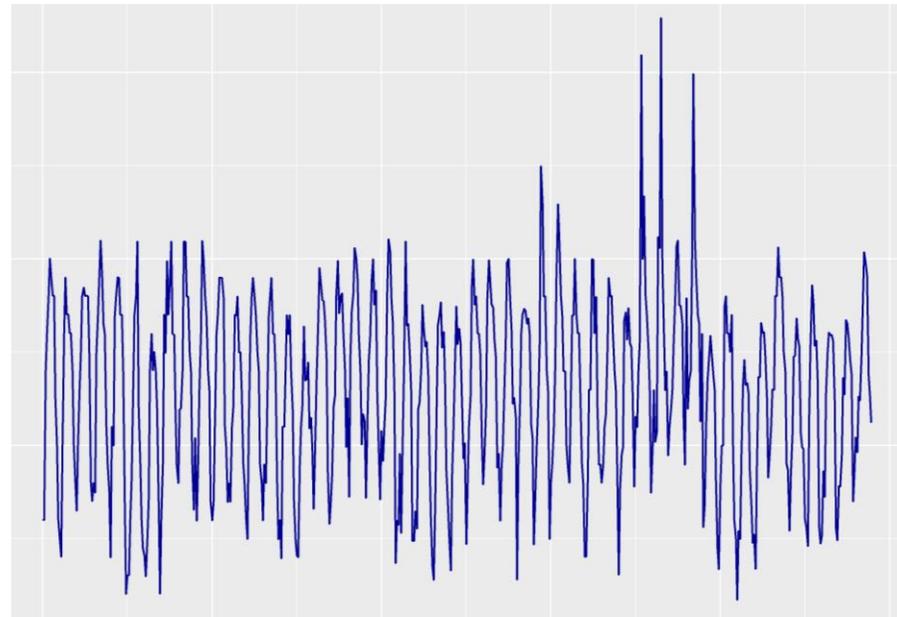
Valores sospechosos



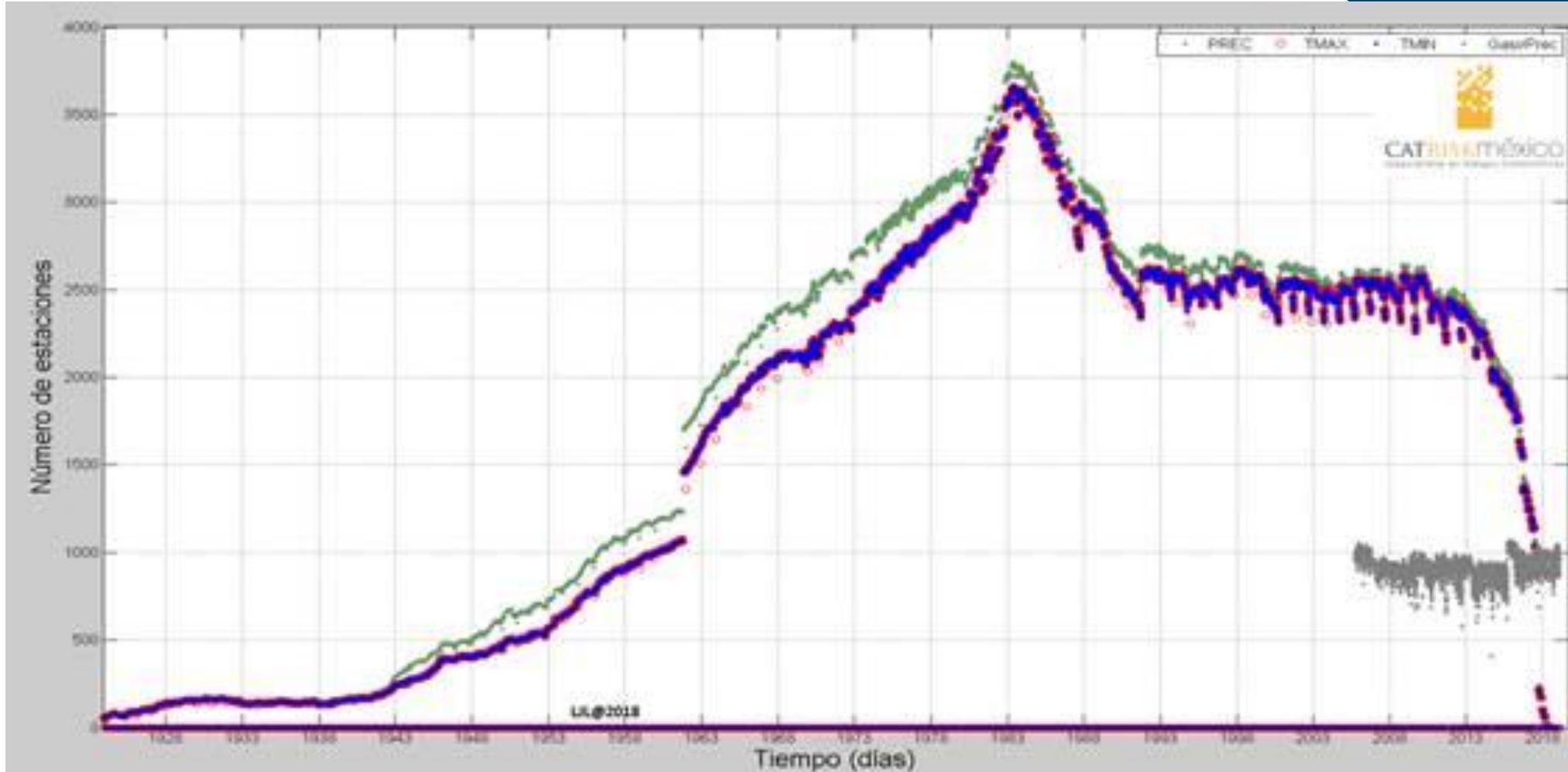
Ausencia de datos



Comportamiento sospechoso de la serie



Registros históricos de estaciones meteorológicas en México



Gráfica proporcionada por la Mtra. Laura Jiménez Lagunes de CatRisk México.

1.2. Antecedentes de herramientas para la homogeneización de series de observaciones meteorológicas

- Existen varios paquetes que han implementando métodos de control de calidad utilizados por la comunidad climatológica.

<http://www.climatol.eu/tt-hom/index.html>

- La acción COST Es0601 (*Advances in homogenisation methods of climate series: an integrated approach, "HOME"*), financió la comparación de los diferentes paquetes de programación (Venema et al., 2012).

- Proyecto MLULTITEST realizó una comparación de los métodos actualizados que pudieran ejecutarse automáticamente.

<http://www.climatol.eu/MULTITEST/>

- INDECIS es un proyecto europeo cuyo objetivo es aplicar métodos en series diarias de diversas variables climáticas.

Guijarro, 2018

Paquete	Resolución de tiempo	Método de detección	Selección de series de referencia	Estadística de detección	Variables Climáticas	Tolerancia de datos faltantes
ACMANT	Mensual y diaria	Referencia	Correlación	Caussinus-Lyazrhi	Temperatura y precipitación	Muy alta
Climatol	Mensual y diaria	Referencia	Distancia	SNHT	Cualquiera	Muy alta
MASH	Mensual y diaria	Múltiples referencias	Correlación	MLR y prueba de hipótesis	Cualquiera	30 %
RHTest	Mensual y diaria	Referencia	Correlación	F test	Cualquiera	–
USHCN	Mensual	Por pares	Correlación	MLR	Temperatura	Muy alto
HOMER	Mensual	Por pares	Correlación	Probabilidad penalizada	Cualquiera	15 años

Métodos del proyecto MULTITEST (OPACE2 y WMO Commission for Climatology (2017);Pita, 2018).

1.3. Paquete de R *Climatol*

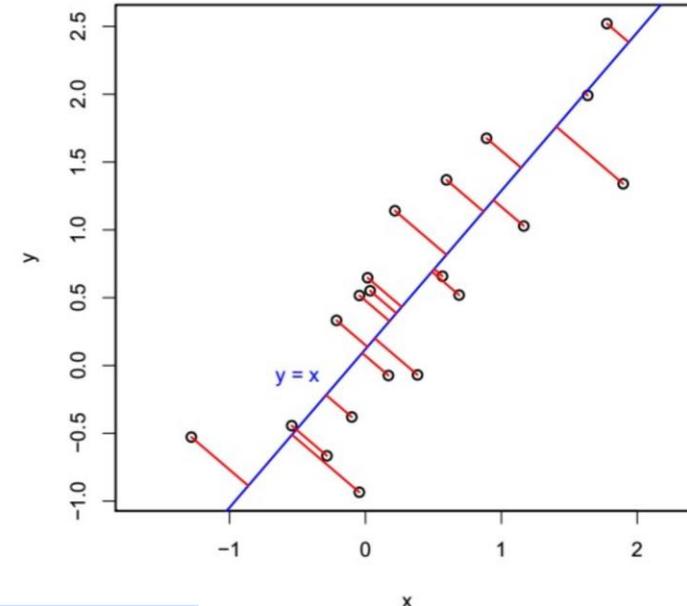
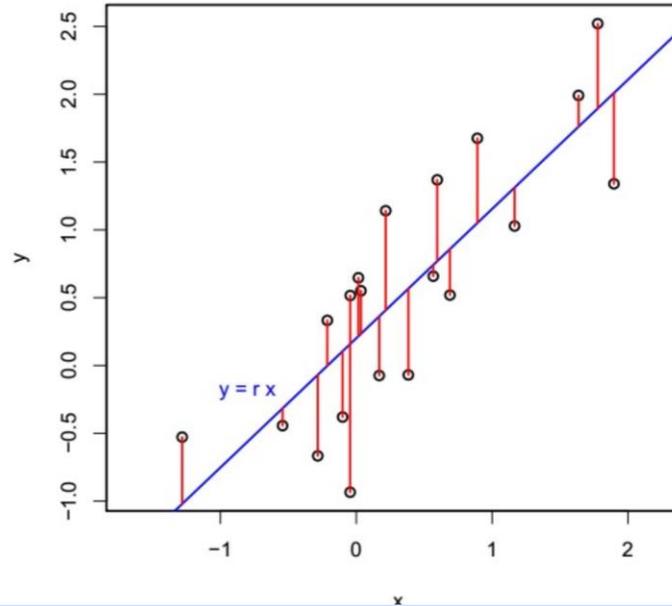
- ❖ Control de calidad
- ❖ Homogeneización
- ❖ Relleno de datos ausentes

Metodología

Basándose en el relleno de datos ausentes; en sus inicios estaba enfocado a rellenar mediante estimaciones calculadas a partir de las series mas próximas. Se adapta el método de Paulhus y Kohler (1952) para rellenar precipitaciones diarias mediante promedios de valores cercanos (Guijarro, 2018).

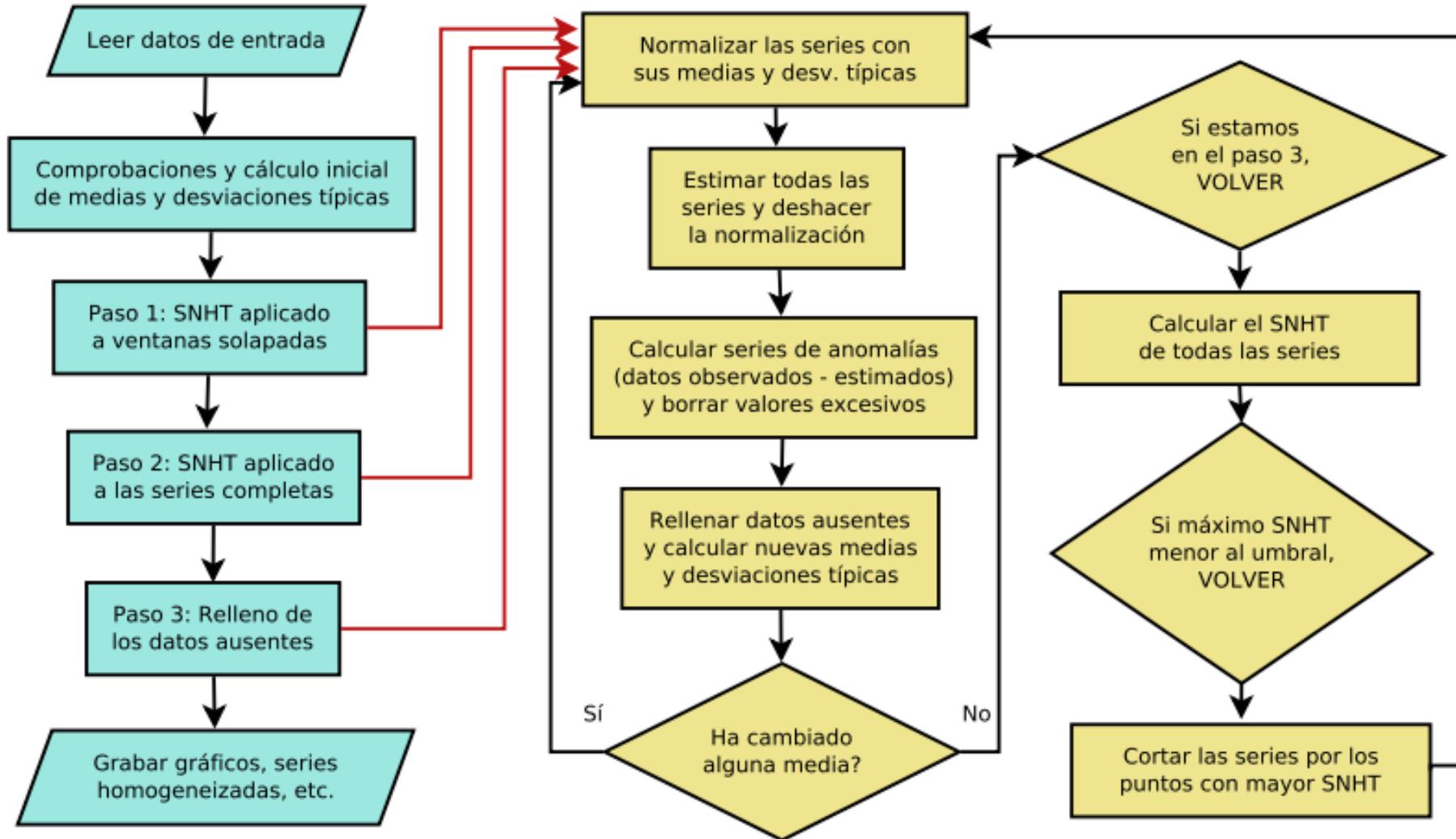
1. Restar la media: $x = X - m_X$
2. Dividir por la media: $x = X/m_X$
3. Estandarizar: $x = (X - m_X)/s_X$

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



La **series de anomalías** permiten:

- Controlar las series y eliminar las anomalías que superen un umbral predefinido.
- Comprobar su homogeneidad a través de la aplicación *Standard Normal Homogeneity Test (SNHT)* (Alexanderson, 1986).



2. Reconstrucción de series climáticas con Climatol



Archivos de entrada

Climatol necesita 2 archivos:

.EST (estaciones)

.DAT (Datos)

¿Cómo obtengo estos archivos?

db2dat: OBDC

daily2climatol: archivos individuales

rclimindex2climatol: formato RClimDex

1	year	mes	dia	TMIN	TMAX	TMED	RANGO
2	2020	3	1	1	29	15	28
3	2020	3	2	0	28	14	28
4	2020	3	3	1	26	13.5	25
5	2020	3	4	4	25	14.5	21
6	2020	3	5	3	26	14.5	23
7	2020	3	6	4	23	13.5	19
8	2020	3	7	4	26	15	22
9	2020	3	8	5	25	15	20
10	2020	3	9	4	28	16	24
11	2020	3	10	4	28	16	24
12	2020	3	11	5	30	17.5	25
13	2020	3	12	6	31	18.5	25
14	2020	3	13	4	30	17	26
15	2020	3	14	5	31	18	26
16	2020	3	15	5	30	17.5	25
17	2020	3	16	9	25	17	16
18	2020	3	17	5	28	16.5	23
19	2020	3	18	5	28	16.5	23
20	2020	3	19	4	29	16.5	25
21	2020	3	20	3	31	17	28
22	2020	3	21	6	29	17.5	23
23	2020	3	22	7	32	19.5	25



30 30.5 29 30.5 30 28.5 30 28.5 30 29.5
30 28.5 29 28.5 29 29.5 29 29.5 29 27.5
29 29.5 29 29.5 28 30.5 30 27.5 26 27.5
28 28.5 29 29.5 30 31.5 30 30.5 30 31.5
28 28.5 27 28.5 28 29.5 29 30.5 31 30.5
29 29.5 29 29.5 30 29.5 30 28.5 28 28.5
29 28.5 29 30.5 31 31.5 31 31.5 30 29.5
31 31.5 31 30.5 29 29.5 30 30.5 31 30.5
29 29.5 29 21.5 26 28.5 26 29.5 29 29.5
30 30.5 24 28.5 29 30.5 29 29.5 28 30.5
30 29.5 28 29.5 29 30.5 30 29.5 28 29.5
29 29.5 30 30.5 31 30.5 30 29.5 29 28.5
30 30.5 31 32.5 29 30.5 29 29.5 29 30.5
26 25.5 26 29.5 28 25.5 23 26.5 29 30.5
30 30.5 31 32.5 30 NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA
NA NA NA NA 34 30.5 37 36.5 36 37.5
37 37.5 36 36.5 35 37.5 37 37.5 38 36.5
31 34 35.5 36 34 35.5 35 35.5 35 36.5
36 35.5 32 31.5 32 30.5 32 33.5 36 37.5
36 38.5 36 34.5 37 37.5 35 34.5 34 32.5
31 34.5 33 33.5 33 35.5 33 36.5 35 36.5
37 37.5 37 36.5 37 36.5 37 36.5 35 35.5
32 30.5 33 35.5 34 32.5 34 35.5 34 32.5
29 30.5 29 29.5 28 30.5 29 31.5 30 31.5
31 31.5 28 28.5 29 27.5 28 31.5 29 30.5

.DAT

-106.3463889 26.95527778 1657 "cn02" ""BALLEZA""
-105.5722222 30.99805556 1060 "cn03" ""BANDERAS""
-106.5063889 28.17027778 1640 "cn04" ""BELISARIO DOMINGUEZ""
-105.1666667 27.28333333 1380 "cn05" ""COLONIA BUFALO (SMN)""
-105.25 27.66666667 1250 "cn06" ""CAMARGO (CFE)""
-105.4666667 28.18333333 1170 "cn07" ""C.A.E. DELICIAS""
-105.4941667 28.27166667 1150 "cn08" ""MEOQUI (SMN)""
-107.0569444 27.91694444 2080 "cn09" ""CARICHI""
-105.1713889 27.68555556 1223 "cn10" ""CAMARGO (SMN)""
-106.9 28.4 1935 "cn11" ""CUAUHTEMOC""
-105.4666667 28.2 1170 "cn12" ""DELICIAS (SMN)""
-104.9136111 27.14055556 1380 "cn13" ""JIMENEZ (SMN)""
-106.4869444 31.73944444 1120 "cn14" ""JUAREZ (SMN)""
-105.1736111 27.29944444 1380 "cn15" ""COLONIA BUFALO (DGE)""
-104.9 29.43333333 909 "cn16" ""CUCHILLO PARADO""
-106.1302778 28.67055556 1509 "cn17" ""CHIHUAHUA (OBS)""
-105.4636111 28.19416667 1173 "cn18" ""DELICIAS (DGE)""
-105.3116667 28.97861111 1080 "cn19" ""LUIS L. LEON""
-106.1558333 27.59111111 1700 "cn20" ""EL MAGUEY""
-104.6747222 29.52777778 820 "cn21" ""EL MEZQUITE""
-105.8475 26.85916667 2027 "cn22" ""EL ORO""
-106.1191667 28.61416667 1464 "cn23" ""PRESA EL REJON (SMN)""
-106.2877778 27.57055556 1600 "cn24" ""EL SITIO""
-106.25 27.05 1580 "cn25" ""EL TULE (SMN)""
-106.3883333 26.47416667 2740 "cn26" ""EL VERGEL (SMN)""
-105.3802778 29.16222222 1420 "cn27" ""GUADALUPE""
-105.7291667 26.90555556 1770 "cn28" ""PRESA PARRAL""
-104.9166667 27.13333333 1370 "cn29" ""JIMENEZ (DGE)""
-105.4305556 28.42305556 1120 "cn30" ""JULIMES""

.EST

La **función de homogeneización** de Climatol se llama ***homogen()***, y su aplicación más trivial solo requiere tres parámetros:

- Acrónimo de la variable
- Año inicial del periodo de análisis
- Año final del periodo de análisis

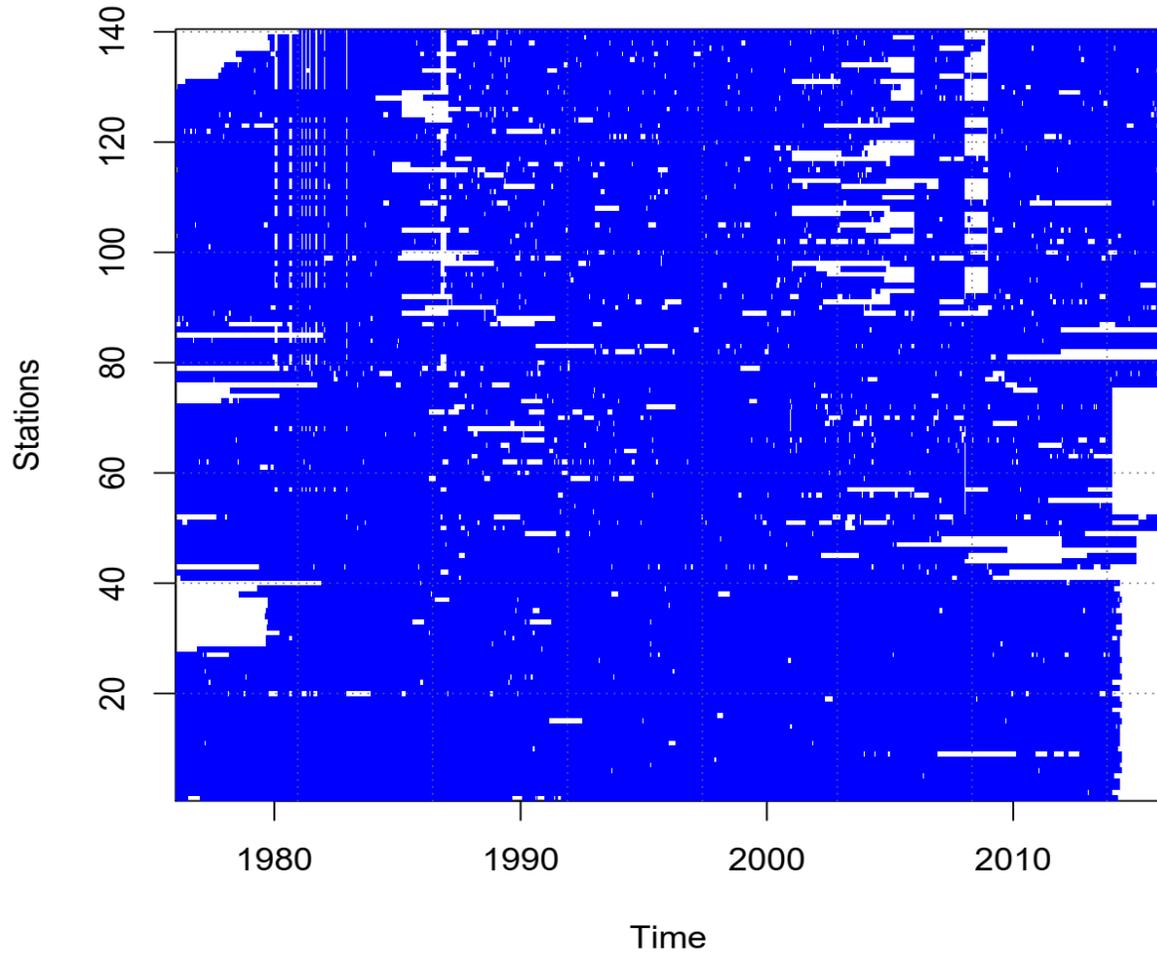
homogen("tmax", 1981, 2015)

Climatol genera un archivo de salida en **formato .pdf**, donde se pueden observar algunos gráficos de diagnóstico.

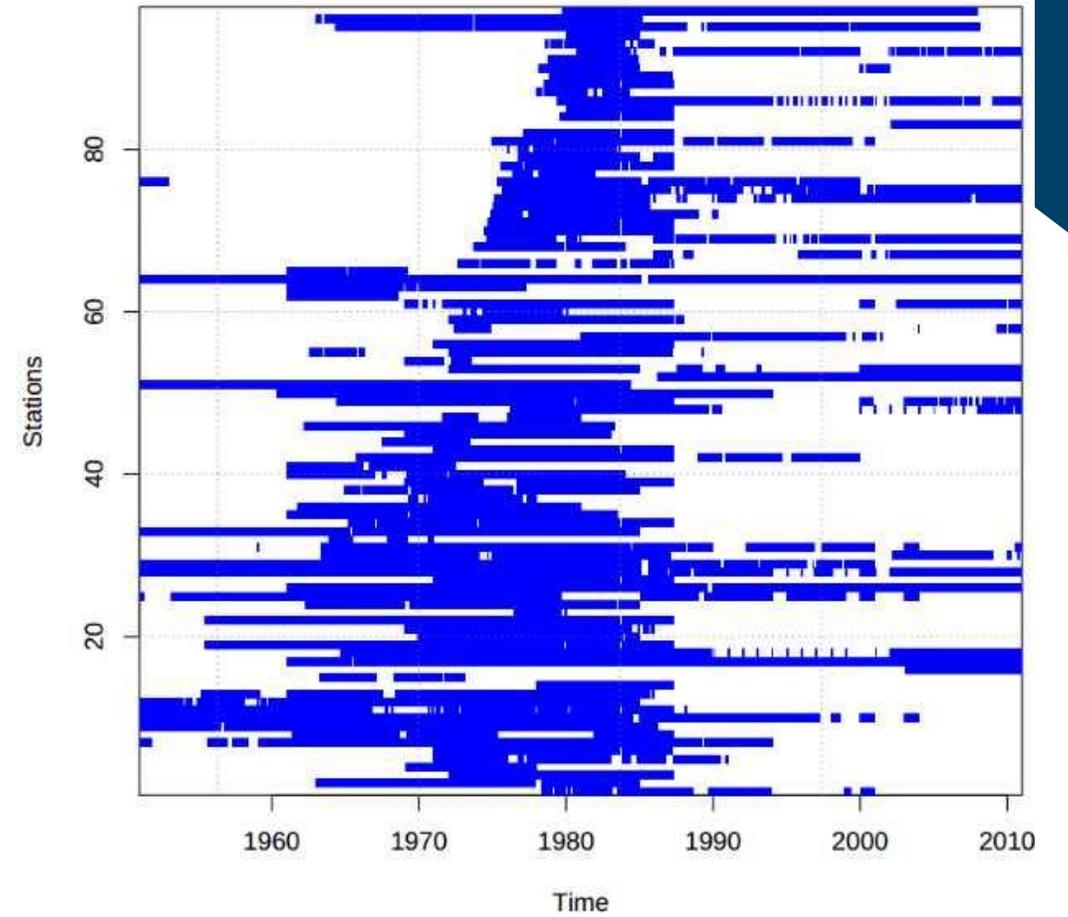
Disponibilidad de datos por estación



tmax data availability



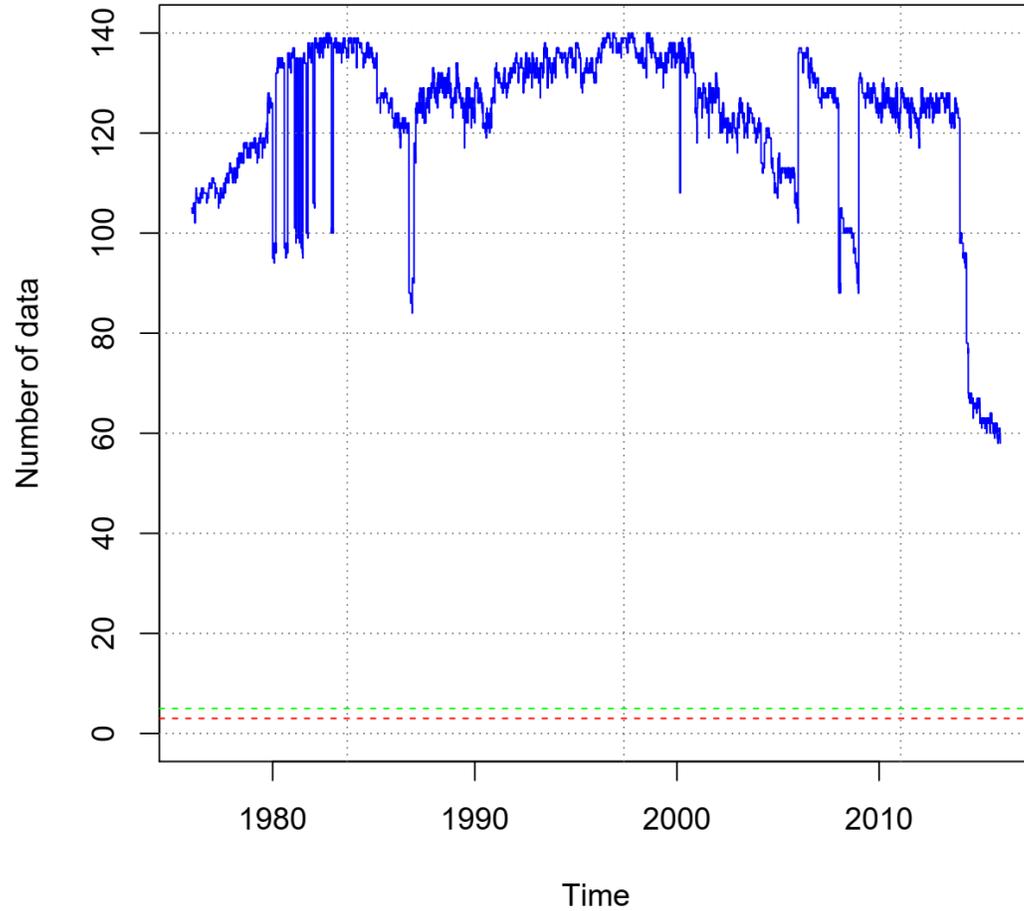
tmax data availability



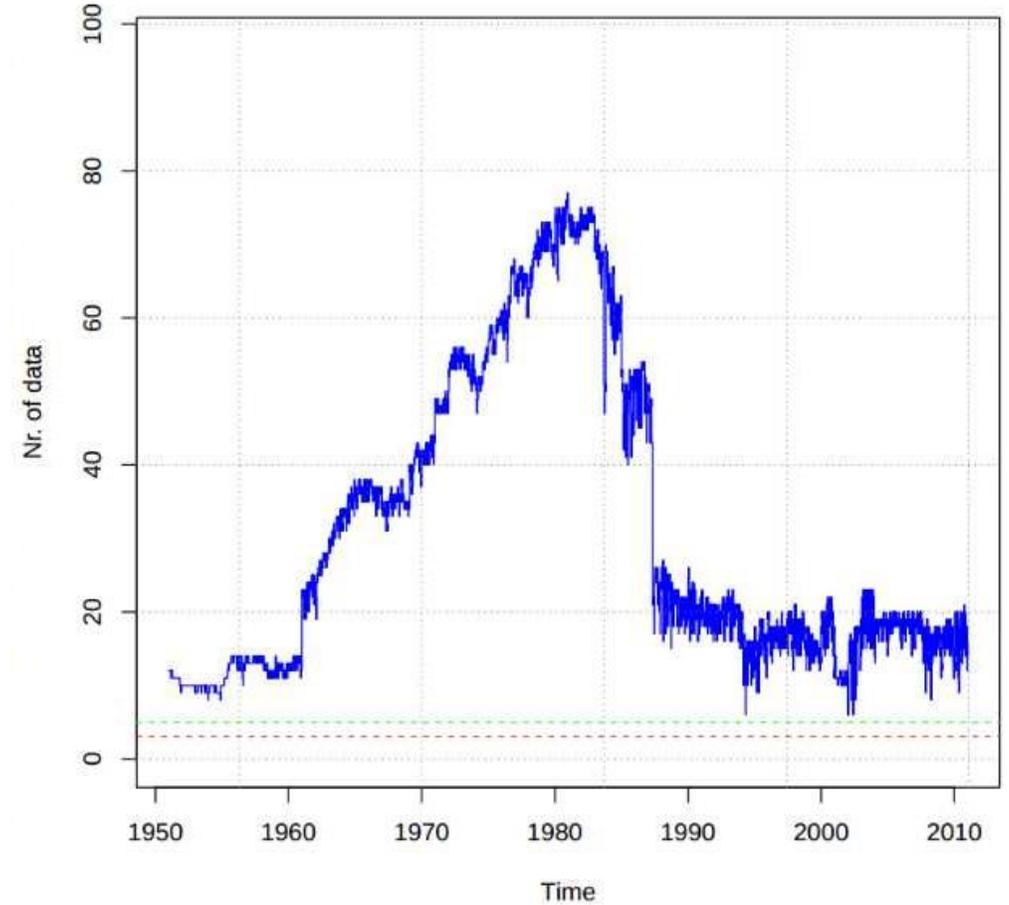
Disponibilidad de datos global



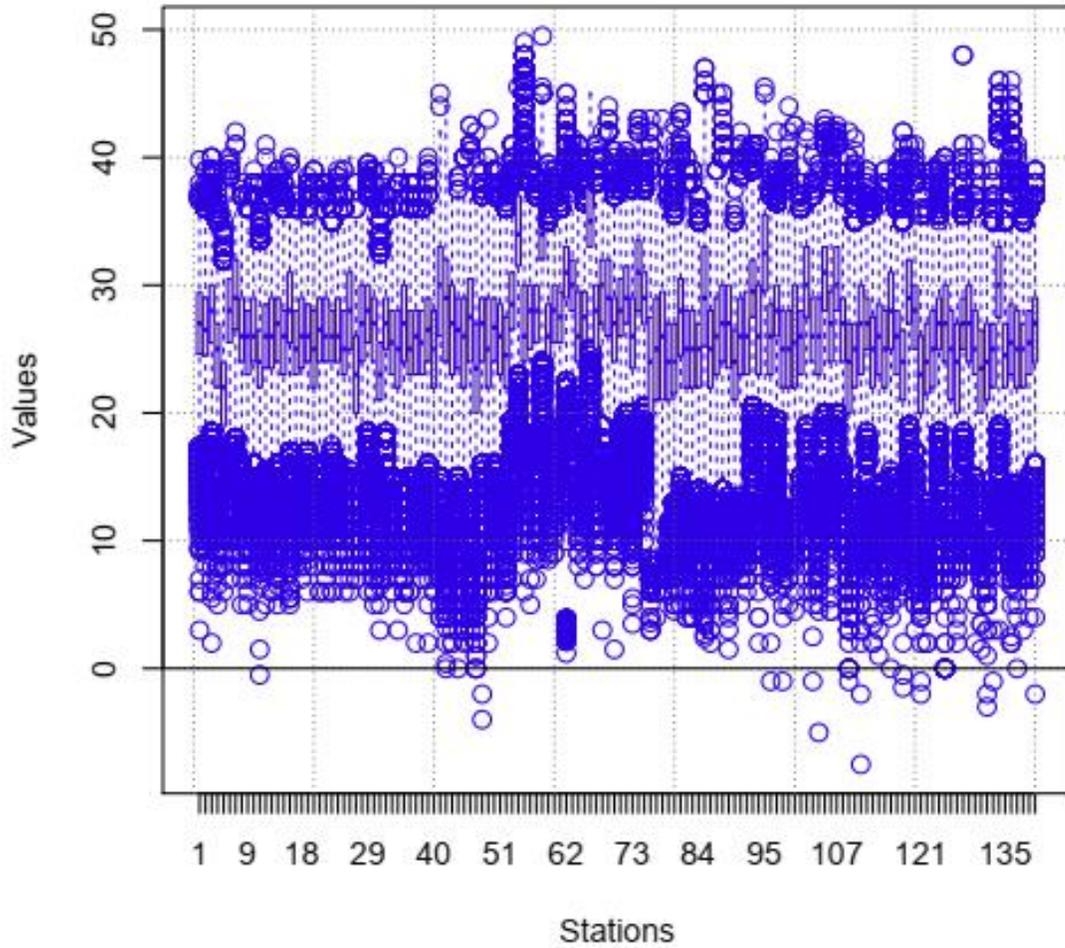
Number of tmax data in all stations



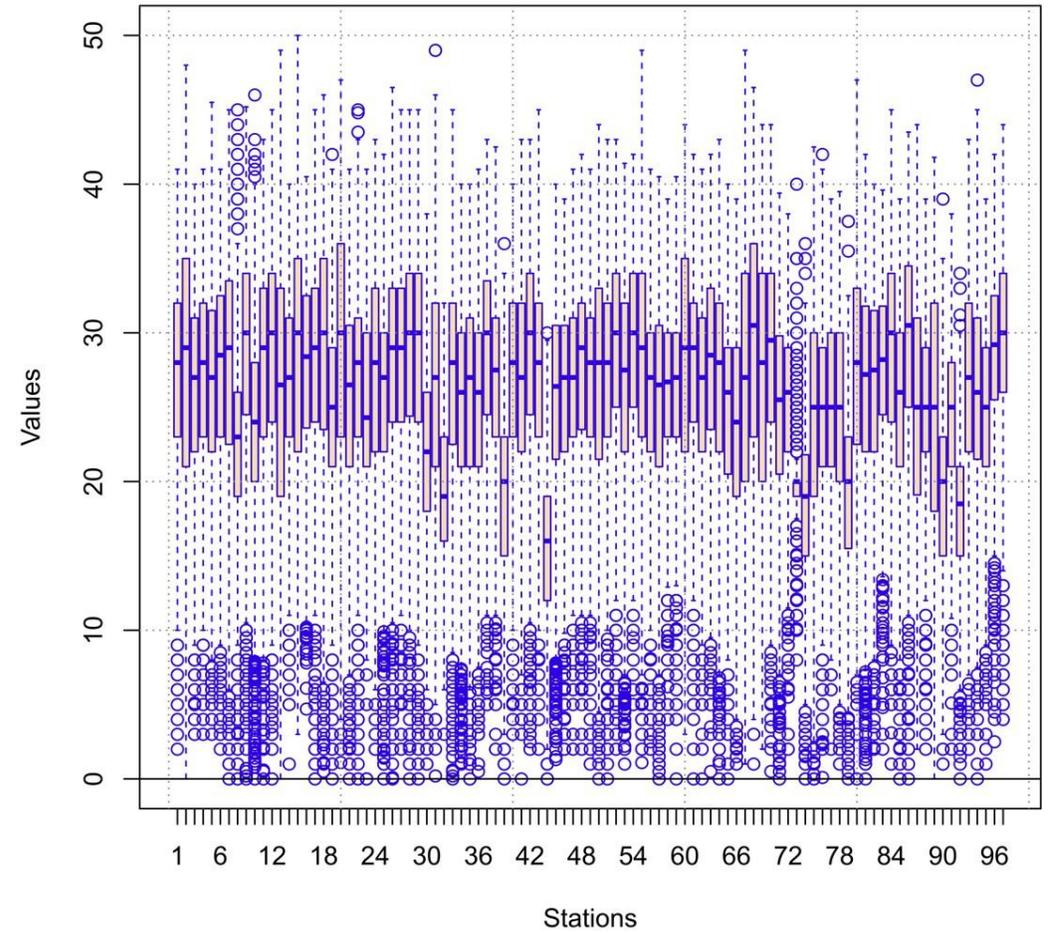
Nr. of tmax data in all stations



Data values of tmax



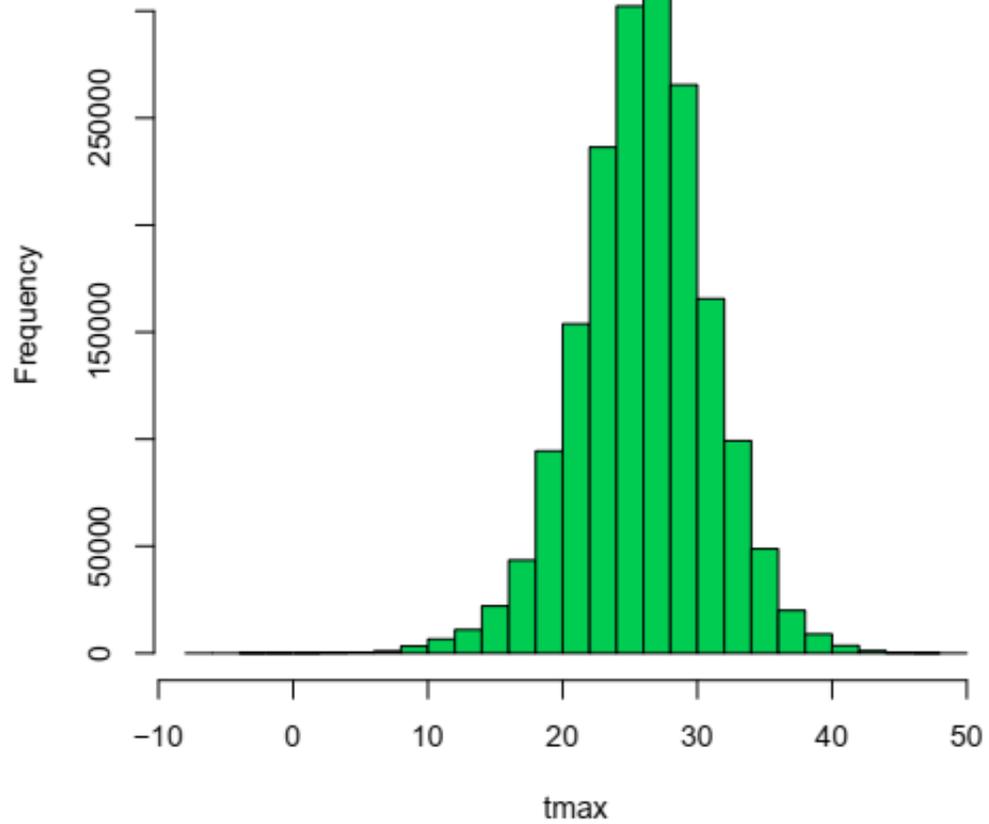
Data values of tmax



Frecuencia de los datos

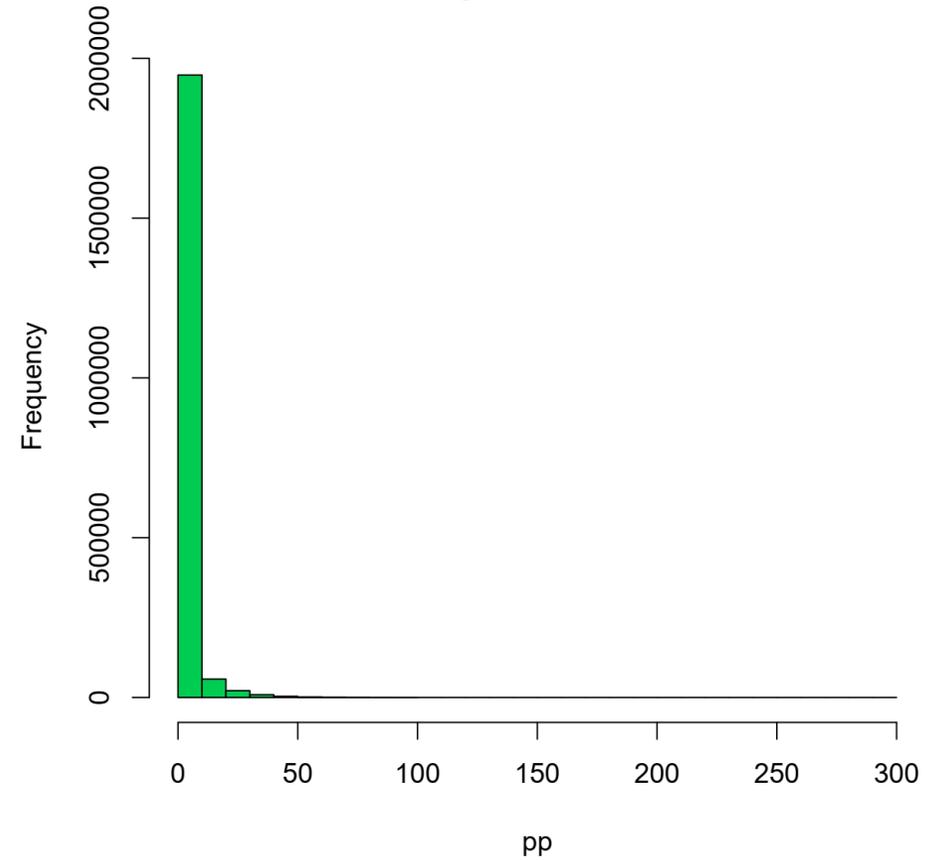


Histogram of all data



Temperatura

Histogram of all data

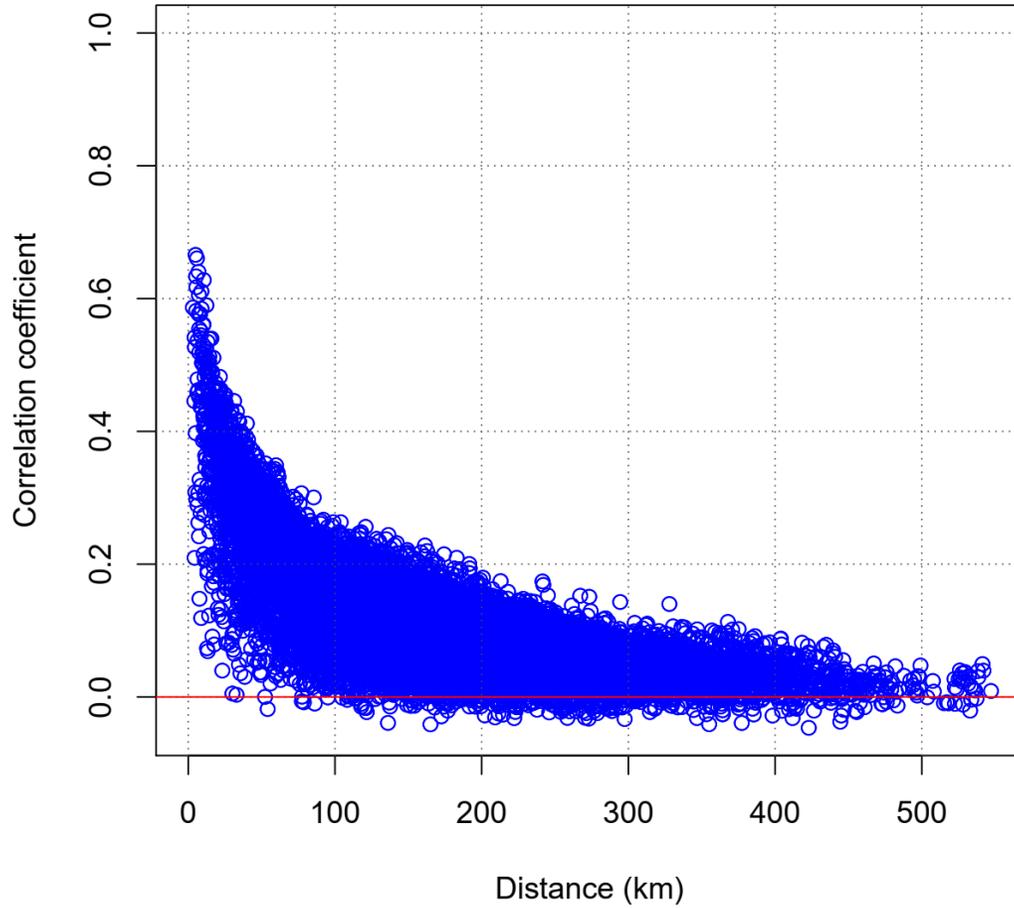


Precipitación

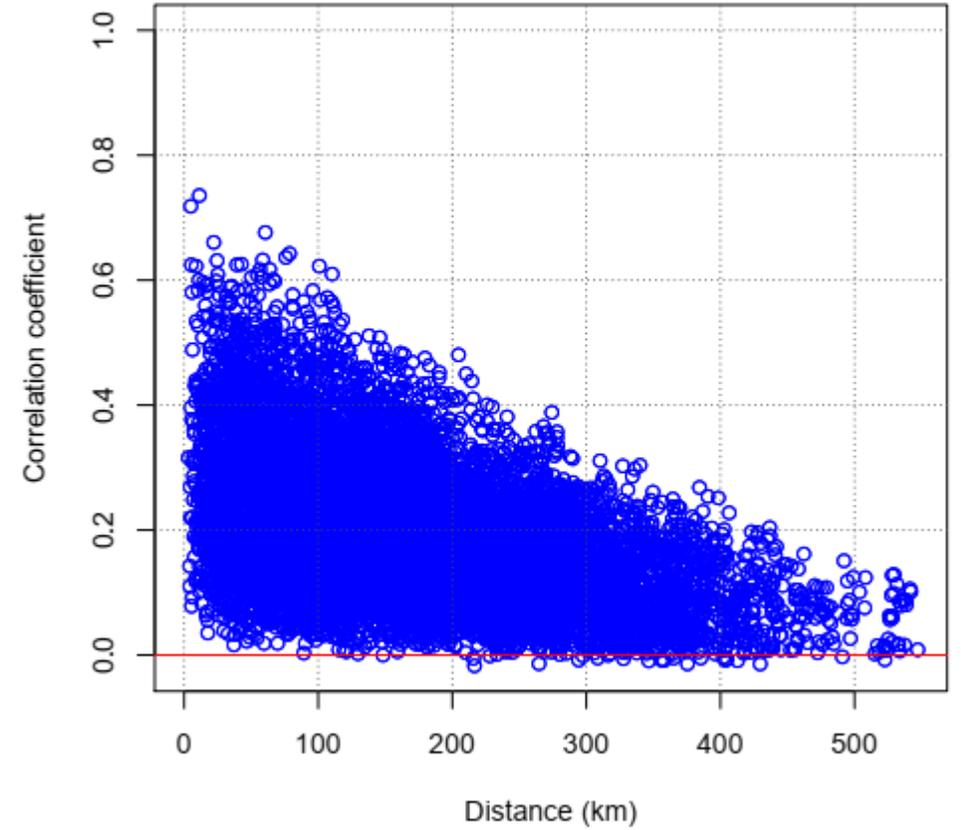
Correlación de la series



Correlogram of the daily series



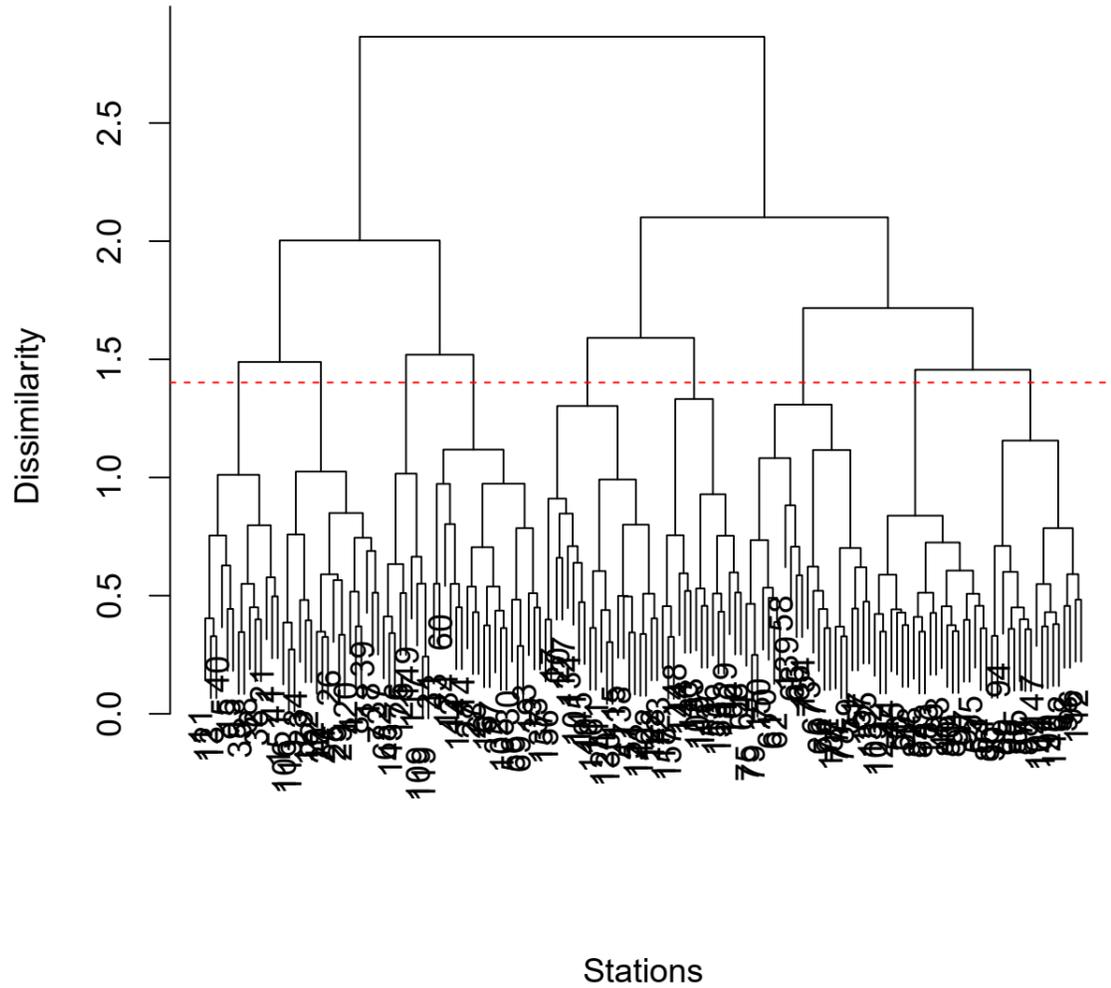
Correlogram of the daily series



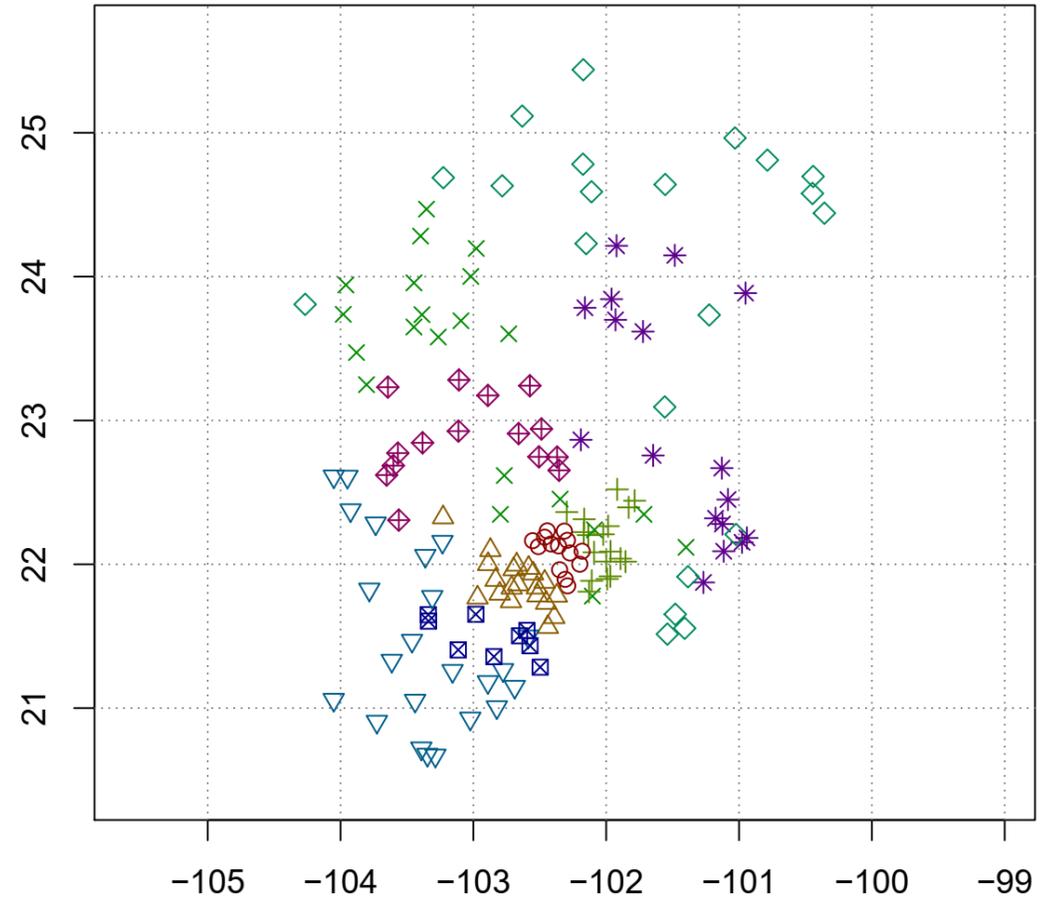
Conglomerado de las series



Dendrogram of station clusters



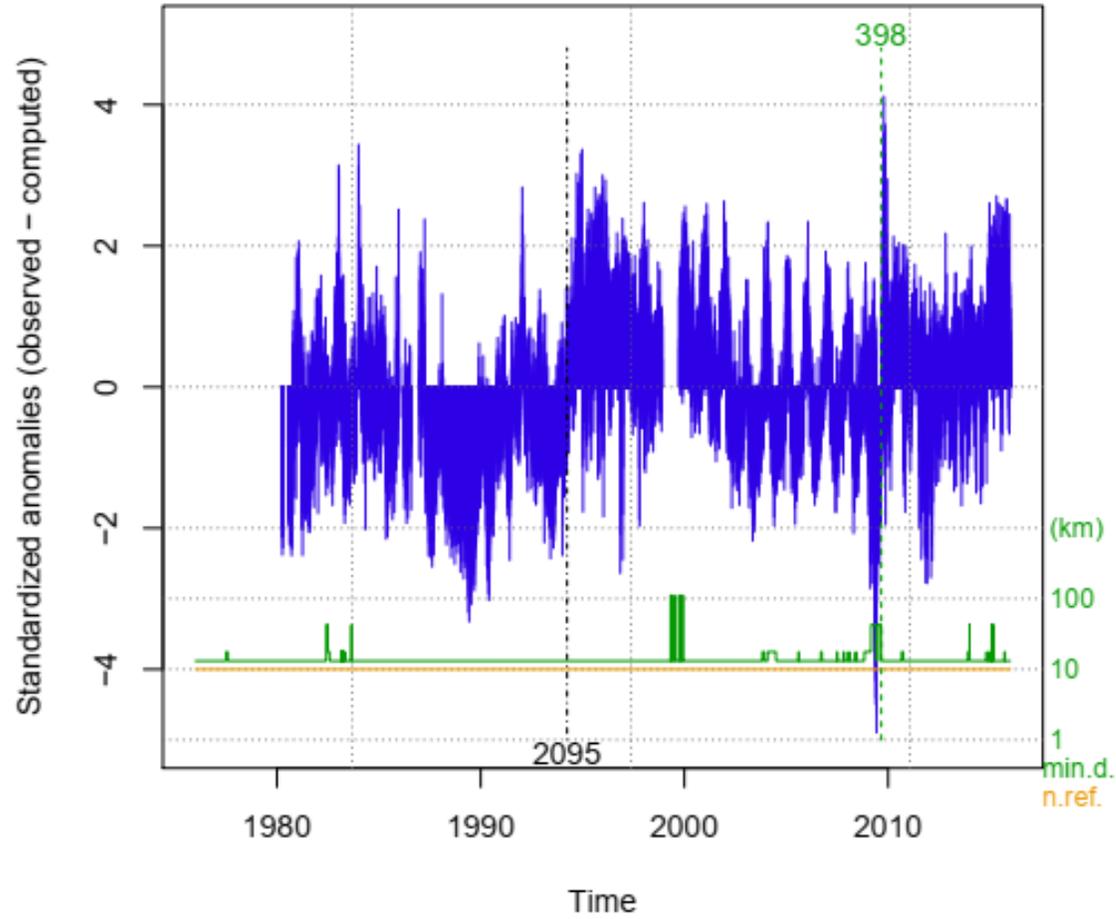
pp station locations (9 clusters)



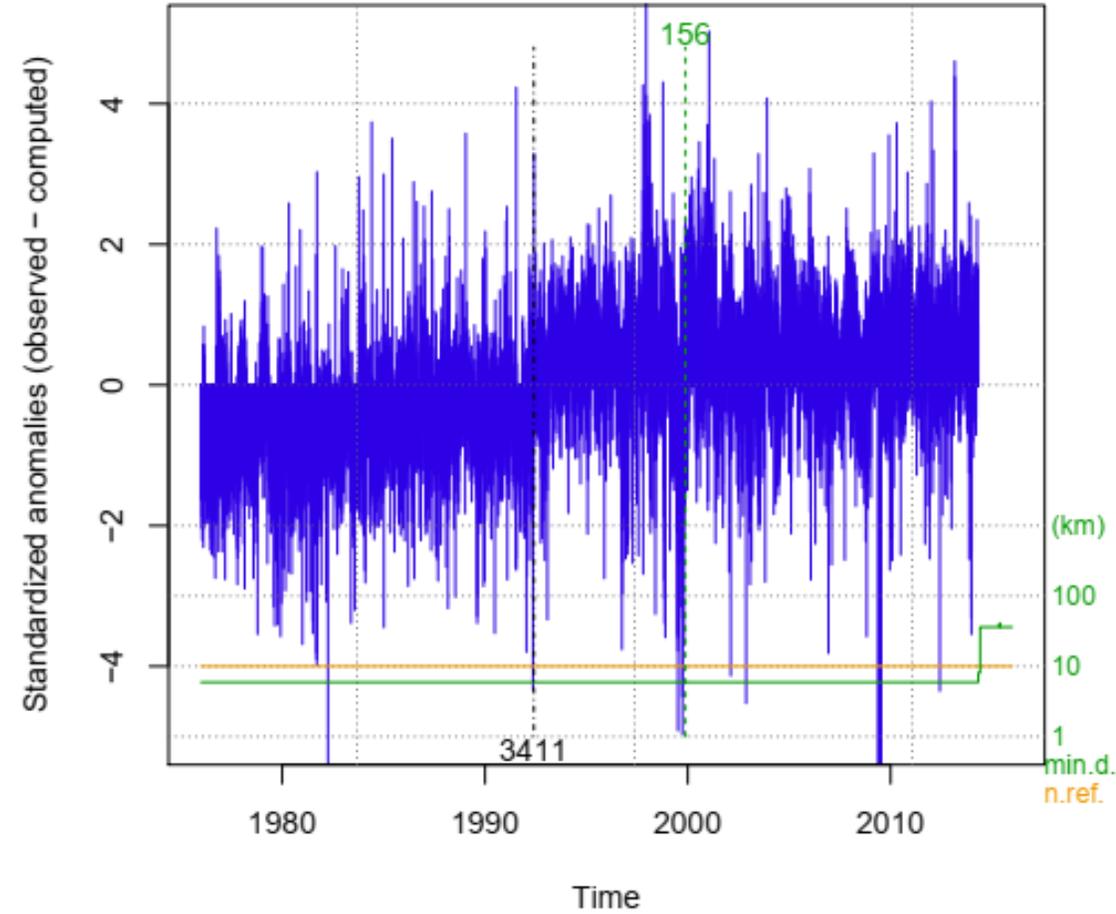
Anomalías estandarizadas



tmax 79 (19137)
SAN JORGE



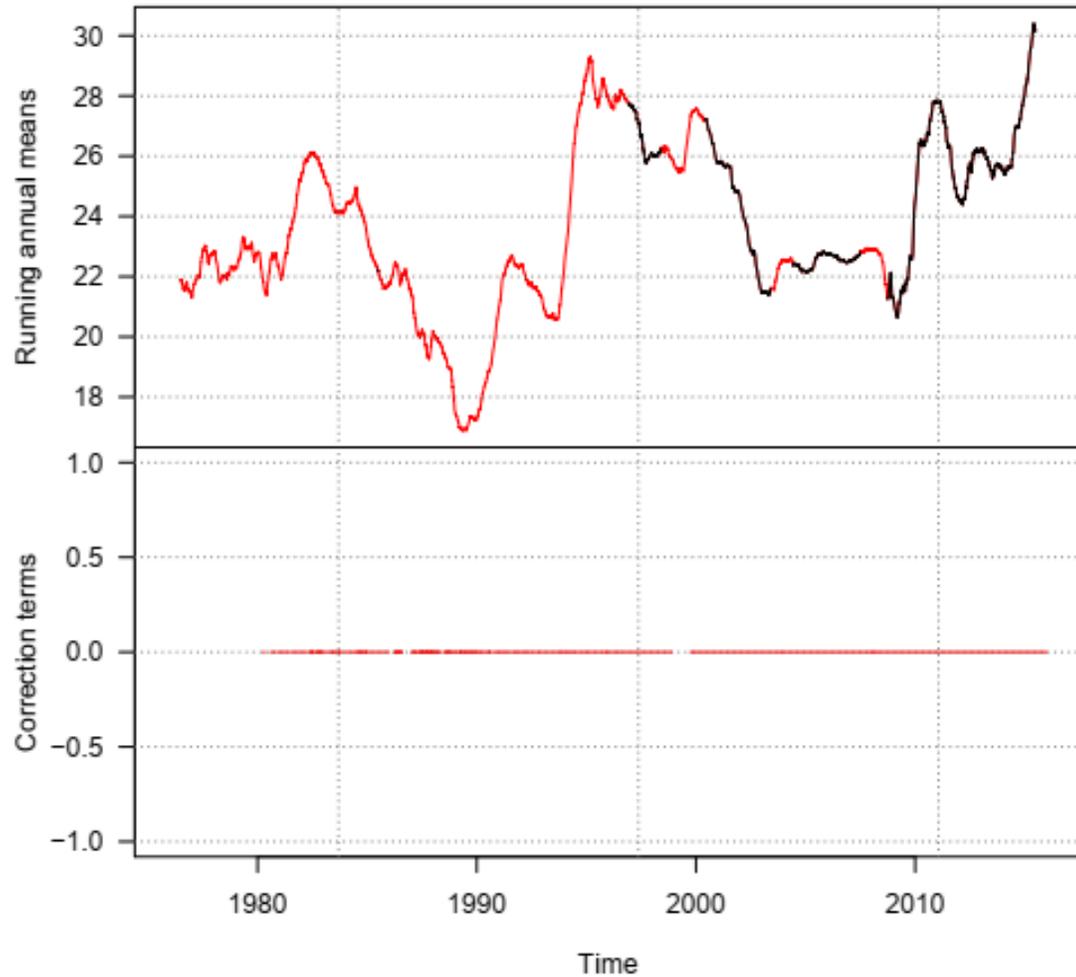
tmax 13 (1019)
PRESA JOCOQUE



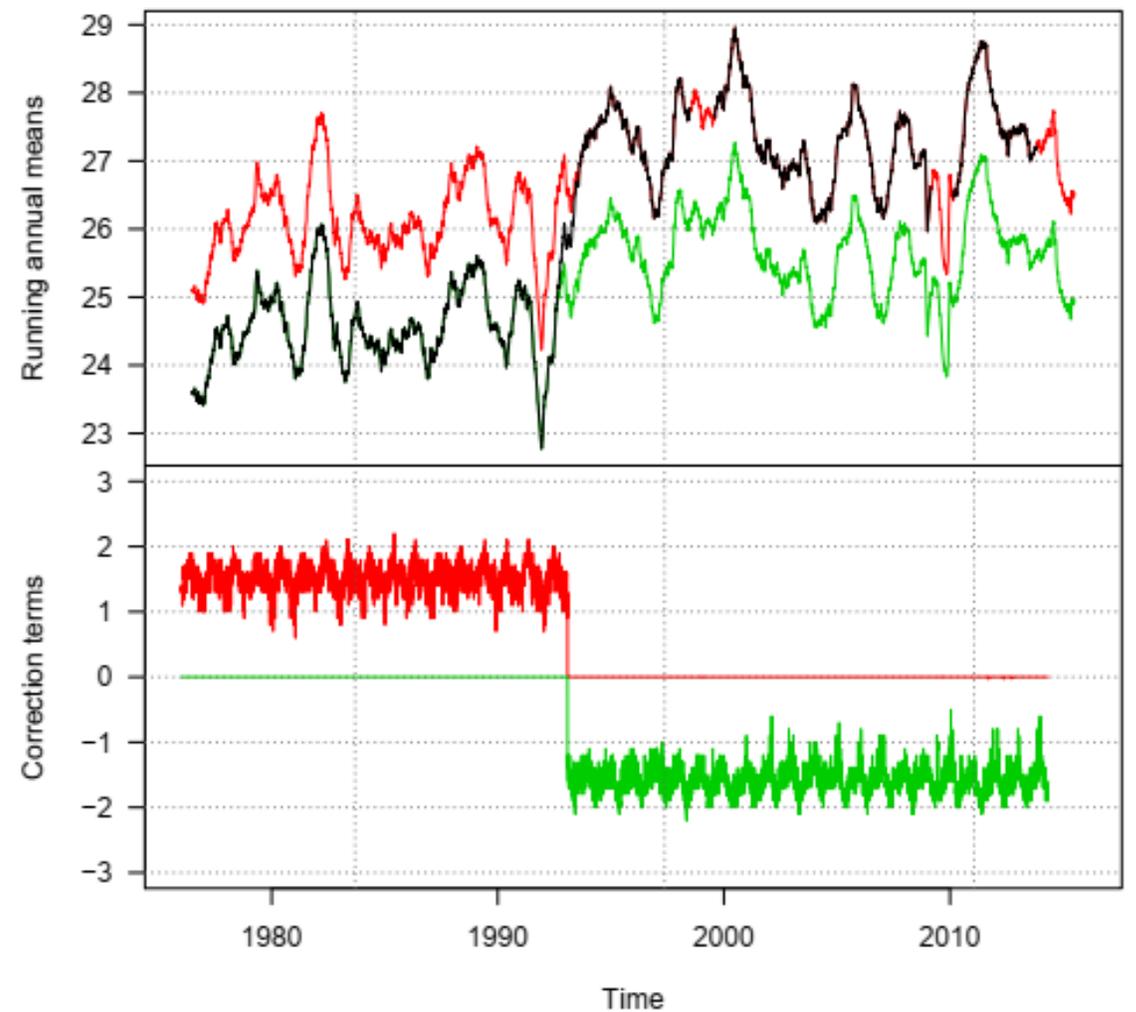
Reconstrucción de series



tmax 79 (19137)
SAN JORGE



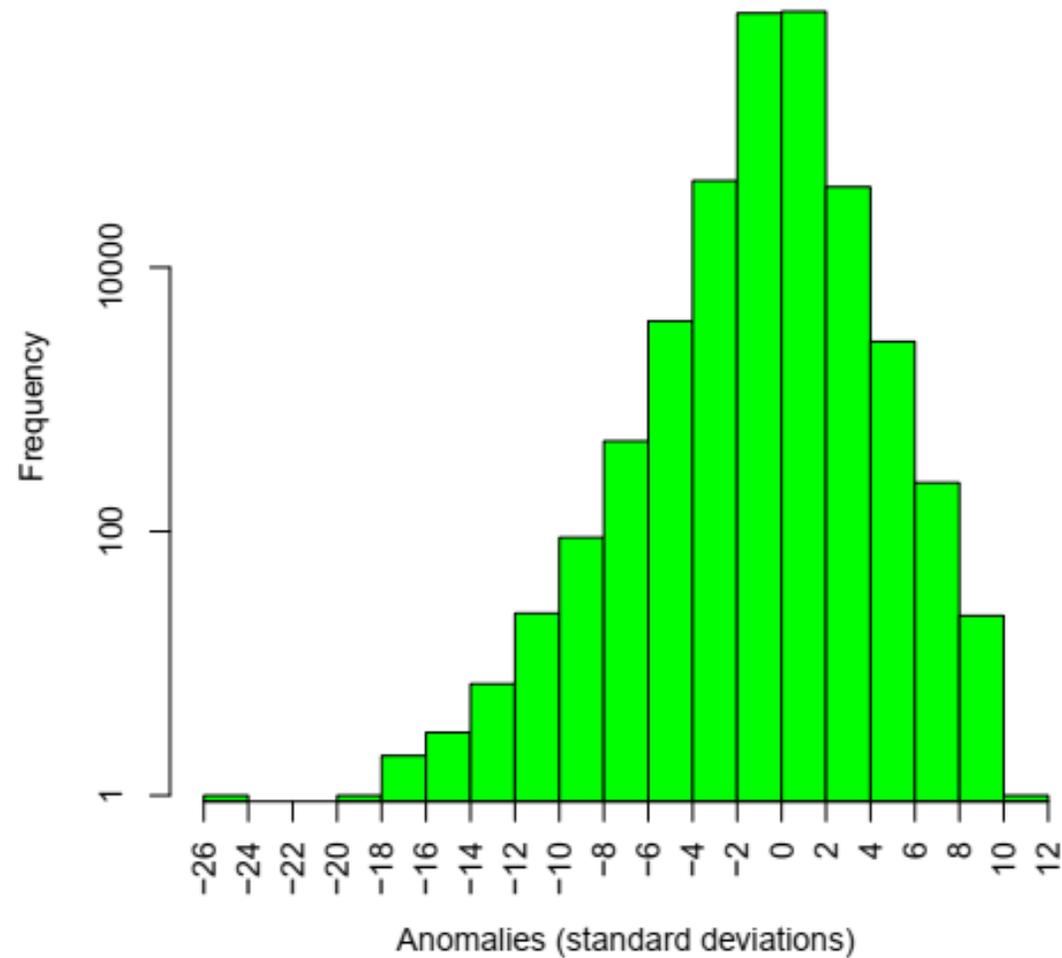
tmax 13 (1019)
PRESA JOCOQUE



Histograma de anomalías



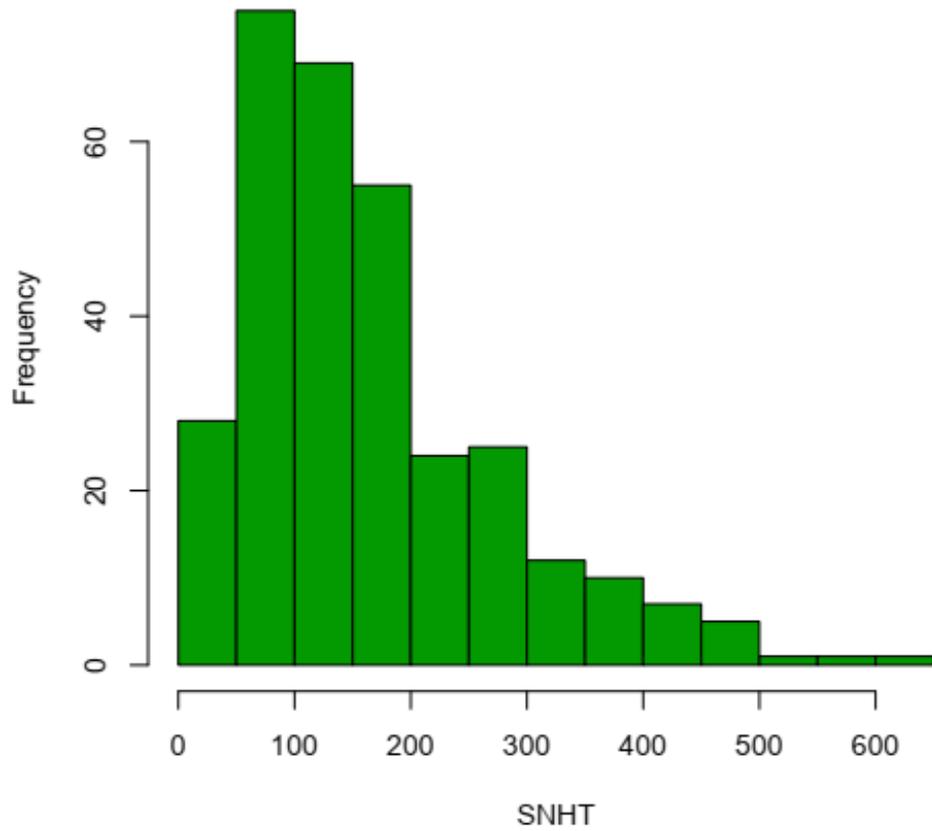
Histogram of standardized anomalies



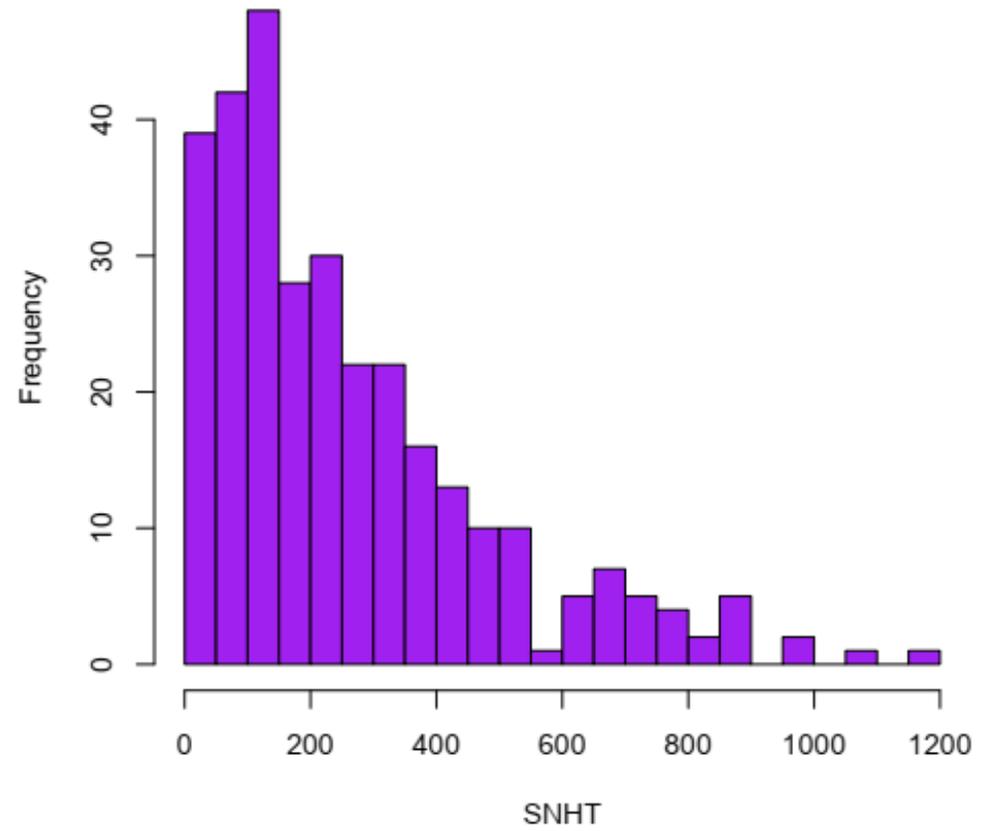
Histogramas SNHT



Histogram of maximum windowed SNHT



Histogram of maximum global SNHT



Ajuste de series diarias usando series mensuales



Función **dd2m**: promedios mensuales y acumulado mensual (precipitación)

- Análisis exploratorio
- Ajuste de series mensuales
- Aplicar: *homogen('test', 1981,2015,metad=TRUE)*

Obtención de series homogeneizadas



Función dahstat()

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2	1951	0	0	7.6	4.8	55.1	125.1	192.6	147.1	245.3	58.2	15.6	0
3	1952	0.5	4.4	1.3	15.8	108	192.4	216.6	220.7	225.1	1.3	55.9	0
4	1953	0	0	1.6	3.1	13.4	129.8	143	218.6	78	88.5	17.3	0
5	1954	0	6.9	6.1	84.7	122.6	175.7	191	117.7	171.6	105.8	16.8	0
6	1955	1.5	0	2.1	0	9	222	337	277	347	66	8	4
7	1956	0	8.5	0	15	116	193.5	145.6	232	180.9	52.5	11.2	0
8	1957	0	6	0	22	43.2	192.8	137.2	179.6	107.4	15.8	7.5	0
9	1958	86.9	14.2	0	24.7	26.2	251.5	181	112.4	315.5	161.1	192.6	43.5
10	1959	9.2	0	0	48.1	59.8	232.4	173.9	267.5	83.8	87.7	3.4	0
11	1960	3.9	0	0	0	52.1	77.8	218	115	124.5	86.7	0	0
12	1961	0	0	5.6	0	29.6	202.6	202	88.4	273.1	30.2	39.2	0
13	1962	0	0	0	15.9	36.2	217.2	88.6	203.3	209.6	81.1	13.8	0
14	1963	0	0	19.3	1.3	80.1	214.9	242.8	100.9	117.1	70.2	0	0
15	1964	14.4	0	0	0	93.1	210.8	148.6	122.7	176.3	29.6	27.3	6.8
16	1965	23.1	14.9	0	31.9	25.3	209.9	114	202.9	121.5	91.7	0	1.5
17	1966	1.4	0	4.7	37	117.2	107.4	150.2	126.6	203	83.5	0	0
18	1967	61.7	0	0	0	132.1	181.2	86.4	231.2	162.9	88.8	0	0
19	1968	0	0	0	23	127.9	177	101.6	95.1	109.9	62.1	0	18
20	1969	17.4	0	4.5	0	8.8	47.8	77	365.5	184.8	53.5	0	0
21	1970	0	0	0	0	3.7	206	177.1	96	163.3	30.7	0	0
22	1971	8.3	0	2.2	3.6	56.4	190.3	219.3	94.5	333	173.7	14	0
23	1972	0	0	3	30.6	33.3	167.2	167.5	67.4	133	7	2	0

Obtención de series homogeneizadas



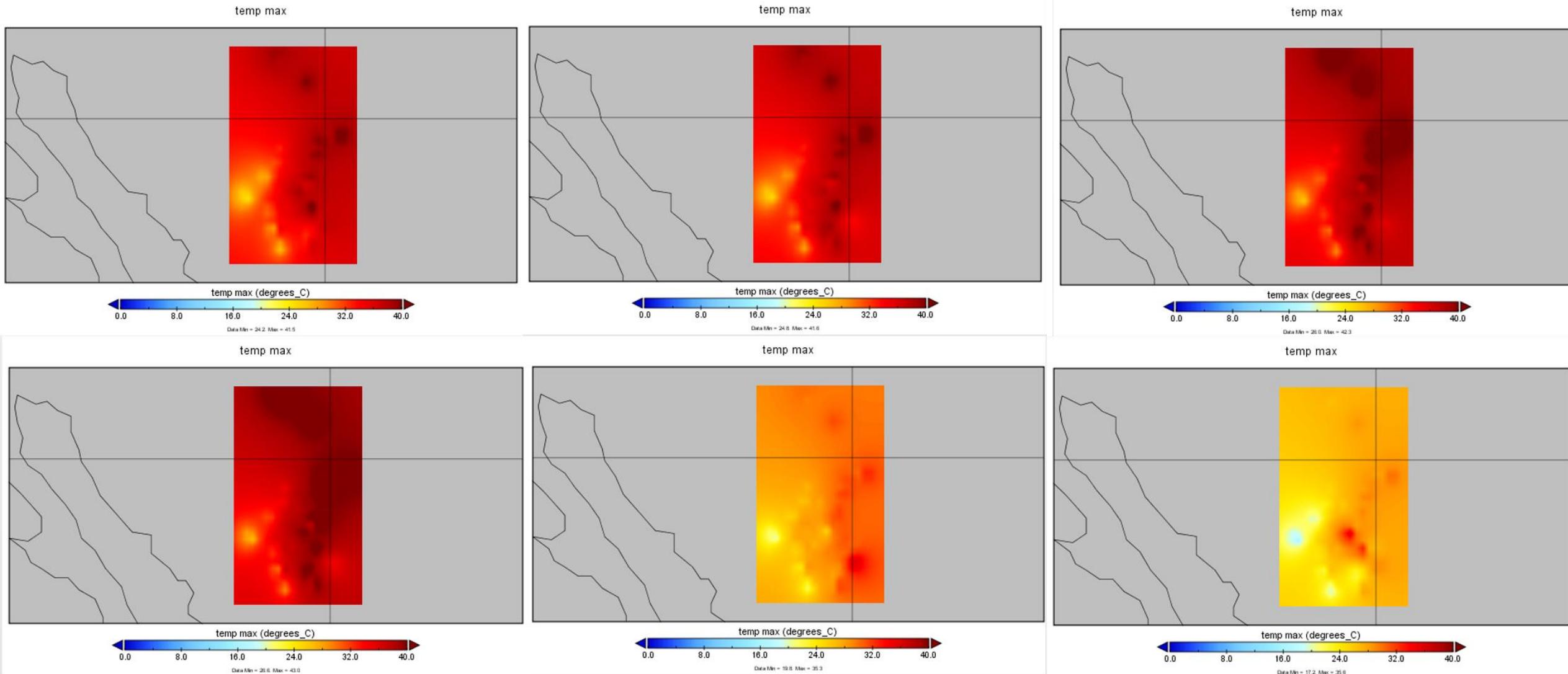
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Date	9002	9019	9020	9022	9051	9067	15053	15173	15222	15228	15232	15242	15256	15354	17002	17004	17006	17013	17014
2	01/01/1951	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	02/01/1951	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	03/01/1951	11.5	12.8	9.4	14.5	3.7	14.6	10.6	8.9	8.3	8.6	7.6	10.5	12.3	10.5	16.8	11.7	8.6	8.3	8.4
5	04/01/1951	5.7	6.1	4.6	7.5	1.9	7.1	5	5	4	4.2	3.8	5	6.4	5.3	1.1	6.5	5	5.3	4.8
6	05/01/1951	73.6	81.2	60	94.7	23.5	93.9	67.5	61.3	53.2	55.5	49.6	67.1	81.9	68.7	42.5	79.6	59.8	59.7	58.3
7	06/01/1951	180	200.3	147.5	229.2	57.2	229.9	167.5	144.7	130.3	135.1	121	165.1	196.8	167.1	197.8	188	138.7	135.5	136.2
8	07/01/1951	277.3	308.6	227.1	353.2	88.1	354.5	257.9	223.3	201.4	208.2	186.9	255	304.2	258	284	289.9	214.3	208.2	209.9
9	08/01/1951	201.9	223.1	164.3	259.2	64.2	257.5	185.8	166	146	152.3	136.6	184.6	223.7	188.1	143.5	216.1	161.3	159.9	157.8
10	09/01/1951	357.5	398.3	293.5	454.8	114.4	456.3	333	284.8	258.3	266.9	239.1	327.7	388.3	331.6	465.8	371	271.8	265.5	267.9
11	10/01/1951	83.4	92.4	68.3	106.2	26.8	106	77.1	66.6	59.9	61.8	55.3	75.7	90.2	77	114.2	87	63.7	63	63
12	11/01/1951	24.2	27.2	19.9	30.5	7.6	31	22.8	18.9	17.8	18.2	16.3	22.5	26.2	22.5	35.5	24.4	17.8	16.9	17.5
13	12/01/1951	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	13/01/1951	1	1.2	3	1.2	0.4	1.3	1.1	0.7	0.7	0.7	0.6	1	0.9	0.9	0	0.9	0.6	0.6	0.6
15	14/01/1951	6.2	6.9	2	7.9	1.9	8.1	5.7	5.3	4.7	4.9	4.5	5.9	7.3	5.9	0	6.8	5.2	4.8	4.9
16	15/01/1951	1.8	2	0	2.3	0.5	2.3	1.6	1.6	1.4	1.4	1.3	1.7	2.2	1.7	0	2	1.5	1.4	1.5
17	16/01/1951	21.3	23.1	7.5	27.3	6.8	26.9	19	17.7	15.2	16	14.5	19.2	23.8	19.8	28.3	23.1	17.4	16.9	16.9
18	17/01/1951	147.3	162.4	109.1	189.1	47.1	187.5	135.2	121.3	106.2	110.6	99.1	134.1	162.7	137.2	139.8	157.8	117.6	116.8	115
19	18/01/1951	281	312.8	251.2	356.8	89.8	358.2	262.3	223.7	203.1	209.2	188.2	257.7	304.3	260	333.7	290.7	213.4	208	209.8
20	19/01/1951	299.4	331.2	194.9	384.7	94.4	382.9	274.7	247.3	217.7	227.2	203.5	274.7	334.1	280.1	287.3	320.8	239.1	234.5	233.2
21	20/01/1951	288.6	316.6	132.8	374.9	90.6	369.1	261	246.3	209.8	220.9	198.2	263.9	329.2	271.5	237.7	318.6	241.3	239	233.6
22	21/01/1951	300.3	329.3	187	387.4	95.2	382.2	273.2	250.5	216.3	226.2	202.9	272.9	334.7	280.3	294.7	325.8	244.2	243.6	238.6
23	22/01/1951	2.5	2.8	4.5	2.9	0.8	3.1	2.4	1.8	1.8	1.8	1.5	2.2	2.5	2.2	0	2.3	1.7	1.6	1.6

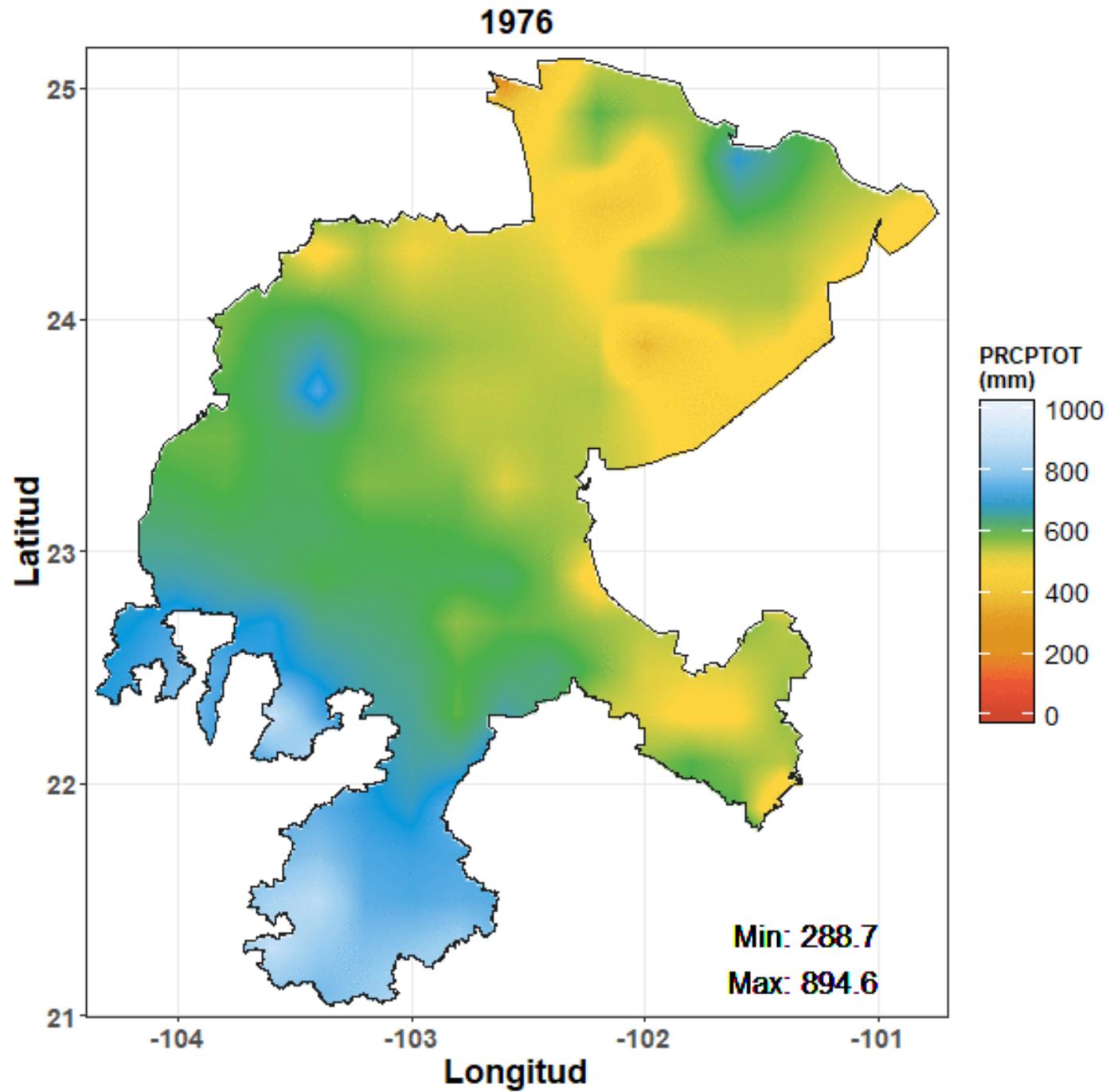
Obtención de series homogeneizadas en rejillas



Función *dahgrid()*

- Valores normalizados
- Uso de Climate Data Operators (CDO) para obtener mallas con valores de temperatura en °C





3. Conclusiones



- Mejorar las estrategias para conservar las estaciones funcionales hasta la actualidad.
- Incentivar los programas de calibración y mantenimiento de estaciones.
- No contar con datos suficientes y de buena calidad provocaría que los análisis climáticos puedan ser poco confiables.

Links de interés



Detección de cambio climático en la cuenca del río Conchos, mediante índices climáticos

<http://repositorio.imta.mx/handle/20.500.12013/2062>

Comparing climate change indices between a northern (arid) and a southern (humid) basin in Mexico during the last decades

<https://www.adv-sci-res.net/15/231/2018/>

Impacto del cambio climático para la gestión integral de la cuenca hidrológica del río Apatlaco, Capítulo III, Homogeneización de datos y cálculo de índices de cambio climático en la cuenca del río Apatlaco.

http://www.imta.gob.mx/biblioteca/libros_html/rio_apatlaco/cambio_climatico_rio_aplatlaco.pdf

Analysis of Anomalies and Trends of Climate Change Indices in Zacatecas, Mexico

<https://doi.org/10.3390/cli8040055>

Sitio web Climatol

<http://www.climatol.eu/>

Documentación de Climatol en el lenguaje R

<https://CRAN.R-project.org/package=climatol>

Guía de usuario Climatol

http://www.climatol.eu/homog_climatol-en.pdf

Muchas gracias

M.C. Óscar Pita-Díaz

Colaborador en la Subcoordinación de
Hidrometeorología del Instituto Mexicano de
Tecnología del Agua

opitadiaz@outlook.com

Para citar esta presentación:

Pita-Díaz, A. 2020. **La homogeneización de datos climáticos como herramienta de análisis de calidad de datos.** Serie de Seminarios Virtuales 2020. Colegio Mexicano de Ingenieros en Irrigación (COMEII). México. 31 pp.

Consulta el portal del COMEII y sus redes sociales:
www.comeii.com y www.riego.mx

