

Homogenización de series de tiempo mensuales de precipitación en R

Néstor Ricardo Bernal Suárez¹, José Daniel Pabón Caicedo²

Primer Simposio de la Red de Investigadores de Ciencias Atmosféricas en Colombia–
AtmosCol2023, Armenia, Quindío, octubre 25 de 2023

Introducción

Cuando se analizan las series de tiempo climáticas, en particular, la de precipitación, constituye un elemento climático (Eslava, 1994; Jaramillo, 2005), en dichas series se pueden detectar heterogeneidades relacionadas con cambios en la ubicación de la estación, posibles fallas en los instrumentos de medición o de transmisión, u otro tipo de errores en los datos. Es por ello que los diversos tipos de usuarios de datos de variables climáticas requieren realizar un análisis encaminado a la calidad de los datos. Así cuando se hace referencia a la homogenización de series de tiempo, lo que se desea es lograr series que posean una coherencia temática, espacial y temporal.

El taller tiene como objetivo presentar la metodología de homogenización de series de tiempo de precipitación, que incluye las siguientes etapas (Bernal et al., 2011; Bernal et al., 2012; Lombana et al., 2018): i) identificación de estaciones meteorológicas vecinas, ii) definir la estación de análisis y estación de referencia, iii) detección de puntos de cambio en el promedio y iv) ajuste de la serie de análisis, precisamente, se le denomina homogenización de las series, siempre y cuando sea necesario realizarlo, pues es probable detectar heterogeneidades que no constituyen errores en los datos, pues las heterogeneidades, pueden detectar datos particulares en la serie o un tramo de la serie que corresponden al efecto de fenómenos meteorológicos o climáticos, por ejemplo, en la escala de la variabilidad climática interanual, la ocurrencia de eventos El Niño o La Niña generan un efecto temporal en la precipitación, que se pueden analizar con estudio de las anomalías (Pabón y Montealegre, 2017), por lo tanto, se puede señalar que sólo se realizará la homogenización siempre que se detecten errores en los datos, en este contexto, el contraste y la comparación de la serie de análisis con otra (s) estaciones de referencia resulta importante, pues ello permite que las series comparadas guarden una coherencia.

Se ilustran tres estudios de caso empleando las series de tiempo mensuales de precipitación de la red hidrometeorológica del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) en Colombia. El primer caso, corresponde a la región de la Sabana de Bogotá, el segundo en la región del Bajo Magdalena, para estos primeros casos se emplea la curva de dobles masas (Barrero & Sabogal, n.d.; Montealegre, 1990) y el tercero en la región del Alto Cauca, para este caso, se emplea como estación de referencia, la serie de tiempo de precipitación del Centro de climatología de precipitación global (GPCC, por sus siglas en inglés) (Schneider et al., 2022) y se ilustra el script

¹ Grupo de Investigación GIIAUD, Ingeniería Ambiental, Universidad Distrital Francisco José de Caldas

² Grupo Tiempo, Clima y Sociedad, Departamento de Geografía, Universidad Nacional de Colombia

en el software R-Studio para ilustrar de forma general la metodología de homogenización. El enfoque metodológico que se presenta tiene similitudes y aportes con el desarrollo de (Mesa et al., 1997) y (Henríquez, 2012); (Martínez et al., 1996) y (Nieto & Ruiz, 2002).

1. Series de tiempo

Acorde a Wei (2006), “un proceso estocástico es una familia de variables aleatorias indexadas en el tiempo $Z(\omega, t)$, donde ω pertenece a un espacio muestral y t pertenece a un conjunto índice. Para un t fijo es una variable aleatoria. Para un ω dado, $Z(\omega, t)$, como una función de t , es llamada una función muestral o realización” ((Wei S., 2006, pág. 6). Así, por ejemplo, la serie de tiempo de precipitación para la estación ICA-Palmira para el período 1930 a 2020 ubicada en la región del Alto Cauca, es una realización de un proceso estocástico, en este contexto, (Storch & Zwiers, 1999) abordan las series de tiempo para análisis climáticos; de igual forma los aportes de autores en temas referentes a aspectos metodológicos de las series para ciencias atmosféricas (Wilks, 2006; Dale & Hale, 2012).

2. Estructura de autocorrelación

La estructura de autocorrelación de una serie de tiempo se analiza con la función de autocorrelación muestral y función de autocorrelación parcial muestral (Wei S., 2006); la función de autocorrelación muestral para el rezago k , se define como el cociente de la covarianza muestral para el rezago k y covarianza muestral para el rezago 0 (*varianza*):

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}, \quad k = 0, 1, 2, \dots$$

Esta función de autocorrelación permite analizar la estructura de autocorrelación de una serie de tiempo para diferentes rezagos.

3. Pruebas de detección de cambios en el promedio

Para efectos de detectar cambios en el promedio de la serie de tiempo, Barbulescu et al. (2010) ilustra la pruebas de Pettitt (1979) (Barbulescu et al., 2010), otra alternativa es la Prueba de Homogeneidad Normal Estándar (SNHT, por sus siglas en inglés) desarrollada por Alexanderson (1986) (World Meteorological Organization (WMO), 2020).

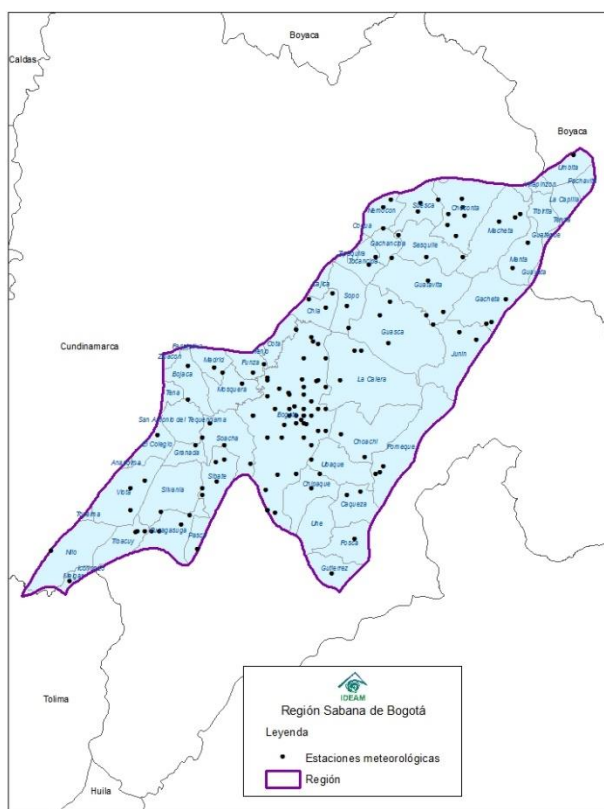
4. Estudios de caso

Estudio de caso 1:

Región climatológica: Sabana de Bogotá

Estación meteorológica de análisis: ESTACIÓN: 2120557, La Primavera, Municipio: Subachoque

Estación meteorológica de referencia (vecina): ESTACIÓN: 2120516, La Ramada, Municipio: Funza



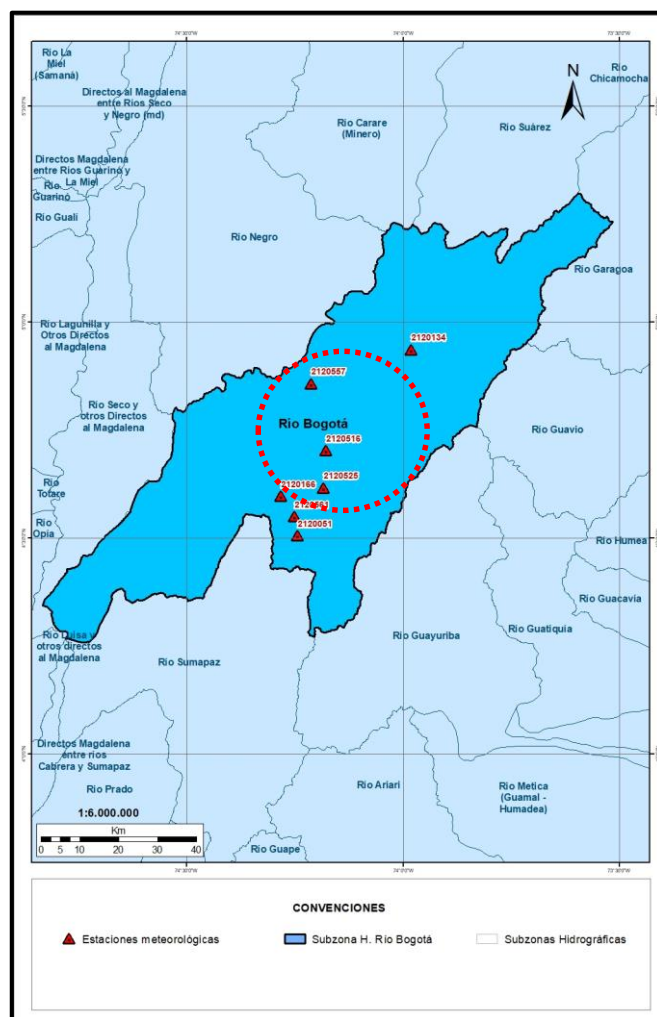
Mapa 1. Estaciones meteorológicas de la región climatológica Sabana de Bogotá
Tomado de Bernal, N., Correa, R. y Rangel, E., 2011

Criterios para identificar una estación meteorológica vecina (Bernal *et al.*, 2012; Lombana *et al.*, 2018)

Una estación “A” es vecina de la estación “B”, si la estación “A” cumple estos cuatro criterios:

- i) Que la estación “A” esté en la misma subzona hidrográfica de la estación “B”
- ii) Que la estación “A” esté en el mismo rango de precipitación anual de la estación “B”
- iii) Que la estación “A” esté en el mismo rango de elevación de la estación “B”
- iv) Que el radio de acción de correlación espacial de precipitación sea significativo, es decir, la distancia entre las distancias A y B no superen un radio de acción específico identificado para cada región climatológica, este radio de acción se identificó empleando el índice de Moran.

Teniendo en cuenta la estructura de correlación espacial evaluada mediante el coeficiente de Moran³, el radio de acción que se estimó para la región Sabana de Bogotá fue de 22,37 km. (Bernal *et al.*, 2011).



Mapa 2. Ejemplo de estaciones meteorológicas de la región climatológica Sabana de Bogotá y radio de acción para identificar estaciones vecinas

³ Para estimar el coeficiente de Moran se empleó como peso de cada par de estaciones, el valor inverso de la distancia entre las dos estaciones y el supuesto de distribución normal empleando el software PASSAGE®.

Estudio de caso 2:

Región Climatológica: Bajo Magdalena

Estación meteorológica de análisis: ESTACIÓN: 2502003, Guacamayo, Municipio: Achí

Estación meteorológica de referencia (vecina): ESTACIÓN: 2502041, La Candelaria, Municipio: San Jacinto del Cauca

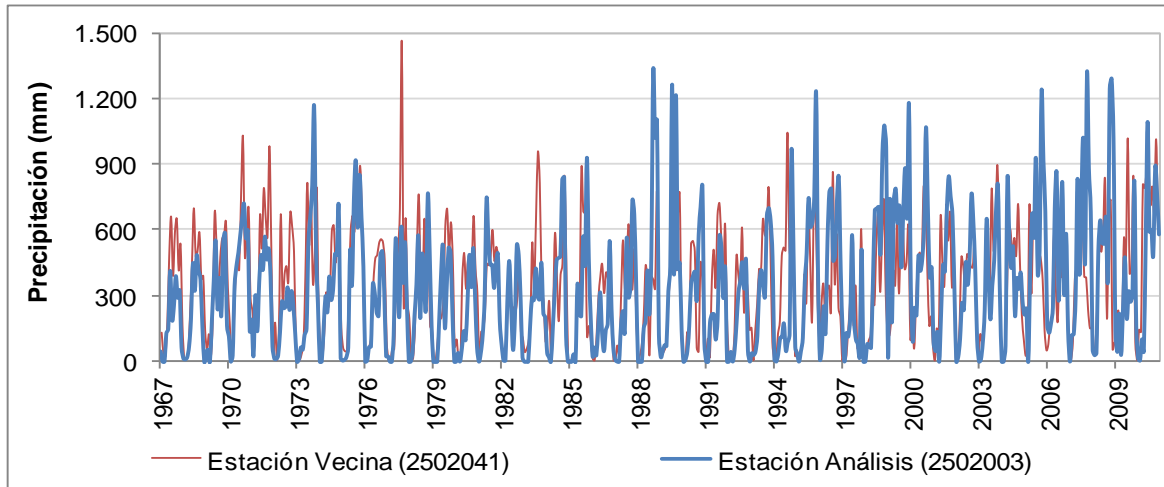


Figura 1. Comparación serie de análisis (2502003) y serie de referencia (2502041)
Tomada de Lombana, L., Bernal, N. y Barrios, J., 2018

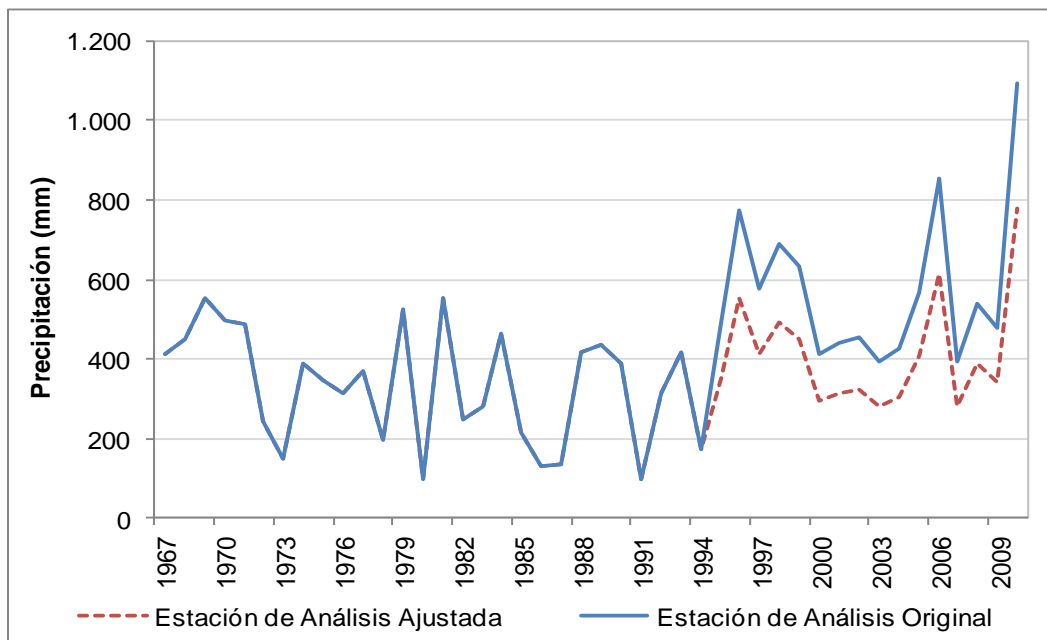


Figura 2. Homogenización mes de junio, Estación 2502003 (Guarandá)
Tomada de Lombana, L., Bernal, N. y Barrios, J., 2018

Estudio de caso 3:

Región Climatológica: Alto Cauca

Estación meteorológica de análisis: ESTACIÓN: 2120516, ICA, Municipio: Palmira

Estación meteorológica de referencia: GPCC

Lectura de datos estación

```
Esta_26075010 <- read.delim("~/Reg_9_Alto_Cauca/26075010.txt", header=FALSE)
```

```
dimnames(Esta_26075010)[[2]] <- c("fecha", "preci_IDEAM")
```

```
Esta_26075010
```

	fecha	preci_IDEAM
1	01/01/1930	42.3
2	01/02/1930	84.2
3	01/03/1930	49.1
4	01/04/1930	180.0
5	01/05/1930	22.1
6	01/06/1930	51.1
7	01/07/1930	0.9
8	01/08/1930	63.6
9	01/09/1930	9.0

lectura fecha

```
fecha_1 <- read.delim("~/Reg_9_Alto_Cauca/fecha5.txt", header=FALSE)
```

```
dimnames(fecha_1)[[2]] <- c("fecha", "secuencia", "ano", "mes")
```

```
fecha_1
```

	fecha	secuencia	ano	mes
1	01/01/1981	1	1981	1
2	01/02/1981	2	1981	2
3	01/03/1981	3	1981	3
4	01/04/1981	4	1981	4
5	01/05/1981	5	1981	5
6	01/06/1981	6	1981	6
7	01/07/1981	7	1981	7
8	01/08/1981	8	1981	8
9	01/09/1981	9	1981	9
10	01/10/1981	10	1981	10
11	01/11/1981	11	1981	11
12	01/12/1981	12	1981	12

unir dos archivos

```
# unir dos archivos 1981 a 2019
```

```
MergedDataset_1 <- merge(fecha_1, Esta_26075010, all=TRUE, by="fecha")
```

```
MergedDataset_1
```

fecha	secuencia	ano	mes	preci_IDEAM	
1	01/01/1981	1	1981	1	6.7
2	01/01/1982	13	1982	1	112.9
3	01/01/1983	25	1983	1	4.2
4	01/01/1984	37	1984	1	94.1
5	01/01/1985	49	1985	1	50.8
6	01/01/1986	61	1986	1	70.7
7	01/01/1987	73	1987	1	15.8
8	01/01/1988	85	1988	1	54.1

```
# ordenar por fecha 1981 a 2019
```

```
Esta_26075010_2 <- with(MergedDataset_1, MergedDataset_1[order(secuencia,  
decreasing=FALSE), ])
```

```
Esta_26075010_2
```

fecha	secuencia	ano	mes	preci_IDEAM	
1	01/01/1981	1	1981	1	6.7
40	01/02/1981	2	1981	2	47.0
79	01/03/1981	3	1981	3	147.0
118	01/04/1981	4	1981	4	258.4
157	01/05/1981	5	1981	5	191.1
196	01/06/1981	6	1981	6	64.2
235	01/07/1981	7	1981	7	17.6
274	01/08/1981	8	1981	8	54.1
313	01/09/1981	9	1981	9	45.4
352	01/10/1981	10	1981	10	120.7
391	01/11/1981	11	1981	11	221.1
430	01/12/1981	12	1981	12	41.8
2	01/01/1982	13	1982	1	112.9
41	01/02/1982	14	1982	2	98.4
80	01/03/1982	15	1982	3	129.9
119	01/04/1982	16	1982	4	210.7
158	01/05/1982	17	1982	5	111.7

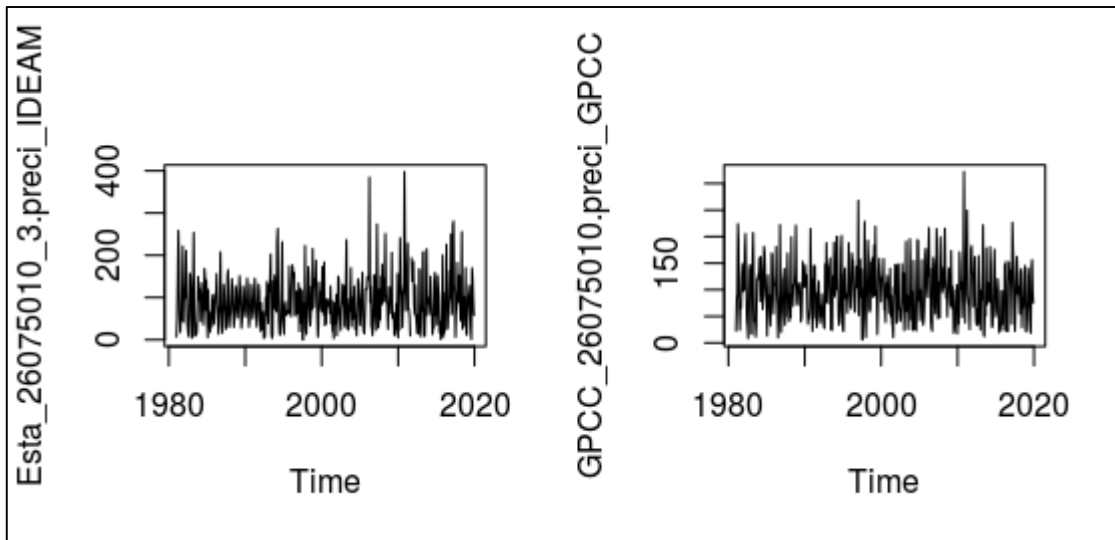


Figura 3. Serie de tiempo de precipitación mensual la estación ICA-Palmira (izquierda) de la red hidrometeorológica, fuente: IDEAM y serie de tiempo de precipitación de GPCC (Schneider et al., 2022).

Correlación:

```
> library(COR)
> cor(x=MergedDataset_IDEAM_GPCC$preci_IDEAM, y=MergedDataset_IDEAM_GPCC$preci_GPCC, use="pairwise
.complete.obs",
+     method=c("pearson"))
[1] 0.8524903
> cor(x=MergedDataset_IDEAM_GPCC$preci_IDEAM, y=MergedDataset_IDEAM_GPCC$preci_GPCC, use="pairwise
.complete.obs",
+     method=c("spearman"))
[1] 0.8760891
> cor(x=MergedDataset_IDEAM_GPCC$preci_IDEAM, y=MergedDataset_IDEAM_GPCC$preci_GPCC, use="pairwise
.complete.obs",
+     method=c("kendall"))
[1] 0.6911169
```

Identificación de datos faltantes:

```
IDEAM_GPCC_26075010_1[369,]
  IDEAM_GPCC_26075010.fecha IDEAM_GPCC_26075010.preci_IDEAM
369              01/09/2011                      NA
  IDEAM_GPCC_26075010.preci_GPCC
369                      88.29
```


Modelo de regresión:

Call:

```
lm(formula = MergedDataset_IDEAM_GPCC$preci_IDEAM ~ MergedDataset_IDEAM_GPCC$preci_GPCC +  
    MergedDataset_IDEAM_GPCC$mes, data = IDEAM_GPCC_26075010)
```

Coefficients:

```
            (Intercept)  
            -0.6025  
MergedDataset_IDEAM_GPCC$preci_GPCC  
            0.9790  
    MergedDataset_IDEAM_GPCC$mes  
            -0.7956
```

```
> summary(mod1)
```

Call:

```
lm(formula = MergedDataset_IDEAM_GPCC$preci_IDEAM ~ MergedDataset_IDEAM_GPCC$preci_GPCC +  
    MergedDataset_IDEAM_GPCC$mes, data = IDEAM_GPCC_26075010)
```

Residuals:

```
      Min       1Q   Median       3Q      Max  
-127.167  -19.701   -3.417   13.324   176.205
```

Coefficients:

```
              Estimate Std. Error t value  
(Intercept)    -0.60248    3.99370  -0.151  
MergedDataset_IDEAM_GPCC$preci_GPCC  0.97904    0.02776  35.265  
MergedDataset_IDEAM_GPCC$mes    -0.79562    0.43208  -1.841  
              Pr(>|t|)  
(Intercept)      0.8802  
MergedDataset_IDEAM_GPCC$preci_GPCC <2e-16 ***  
MergedDataset_IDEAM_GPCC$mes      0.0662 .  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 32.09 on 464 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.7287, Adjusted R-squared: 0.7276

F-statistic: 623.2 on 2 and 464 DF, p-value: < 2.2e-16

Estimación de dato faltante:

	fecha	preci_GPCC	secuencia	ano	mes	preci_IDEAM	Preci_est_IDEAM_1	Preci_est_IDEAM	IDEAM_GPCC_26075010.1.Preci_est_IDEAM
226	01/06/2011	76.71	366	2011	6	72.6	72.6	72.60000	72.60000
265	01/07/2011	69.27	367	2011	7	54.8	54.8	54.80000	54.80000
304	01/08/2011	25.35	368	2011	8	9.4	9.4	9.40000	9.40000
343	01/09/2011	88.29	369	2011	9	N/A	N/A	78.67891	78.67891
382	01/10/2011	176.27	370	2011	10	193.9	193.9	193.90000	193.90000
421	01/11/2011	182.74	371	2011	11	138.1	138.1	138.10000	138.10000
460	01/12/2011	131.58	372	2011	12	142.1	142.1	142.10000	142.10000
32	01/01/2012	129.91	373	2012	1	184.9	184.9	184.90000	184.90000

Detección del cambio en el promedio (prueba de Pettitt):

Pettitt's test for single change-point detection

data: IDEAM_GPCC_26075010_2\$Preci_est_IDEAM

U* = 3324, p-value = 1

alternative hypothesis: two.sided

sample estimates:

probable change point at time K

297

```

> IDEAM_GPCC_26075010_2[295,]
      fecha preci_GPCC secuencia ano mes preci_IDEAM Preci_est_IDEAM_1
259 01/07/2005    33.72      295 2005    7      19.3      19.3
      Preci_est_IDEAM IDEAM_GPCC_26075010_1.Preci_est_IDEAM
259      19.3      19.3
> IDEAM_GPCC_26075010_2[296,]
      fecha preci_GPCC secuencia ano mes preci_IDEAM Preci_est_IDEAM_1
298 01/08/2005    43.72      296 2005    8      14.4      14.4
      Preci_est_IDEAM IDEAM_GPCC_26075010_1.Preci_est_IDEAM
298      14.4      14.4
> IDEAM_GPCC_26075010_2[297,]
      fecha preci_GPCC secuencia ano mes preci_IDEAM Preci_est_IDEAM_1
337 01/09/2005    72.82      297 2005    9      58.3      58.3
      Preci_est_IDEAM IDEAM_GPCC_26075010_1.Preci_est_IDEAM
337      58.3      58.3
> IDEAM_GPCC_26075010_2[298,]
      fecha preci_GPCC secuencia ano mes preci_IDEAM Preci_est_IDEAM_1
376 01/10/2005   166.25      298 2005   10     119.7     119.7
      Preci_est_IDEAM IDEAM_GPCC_26075010_1.Preci_est_IDEAM
376     119.7     119.7
> IDEAM_GPCC_26075010_2[299,]
      fecha preci_GPCC secuencia ano mes preci_IDEAM Preci_est_IDEAM_1
415 01/11/2005   177.21      299 2005   11     118.8     118.8
      Preci_est_IDEAM IDEAM_GPCC_26075010_1.Preci_est_IDEAM
415     118.8     118.8

```

Detección del cambio en el promedio (prueba de Homogeneidad Normal Estándar):

Standard Normal Homogeneity Test (SNHT)

data: IDEAM_GPCC_26075010_2\$Preci_est_IDEAM

T = 3.8327, n = 468, p-value = 0.7018

alternative hypothesis: true delta is not equal to 0

sample estimates:

probable change point at time K
297

```

> plot(out)
> IDEAM_GPCC_26075010_2[295,]
      fecha preci_GPCC secuencia ano mes preci_IDEAM Preci_est_IDEAM_1
259 01/07/2005    33.72      295 2005    7      19.3      19.3
      Preci_est_IDEAM IDEAM_GPCC_26075010_1.Preci_est_IDEAM
259      19.3      19.3
> IDEAM_GPCC_26075010_2[296,]
      fecha preci_GPCC secuencia ano mes preci_IDEAM Preci_est_IDEAM_1
298 01/08/2005    43.72      296 2005    8      14.4      14.4
      Preci_est_IDEAM IDEAM_GPCC_26075010_1.Preci_est_IDEAM
298      14.4      14.4
> IDEAM_GPCC_26075010_2[297,]
      fecha preci_GPCC secuencia ano mes preci_IDEAM Preci_est_IDEAM_1
337 01/09/2005    72.82      297 2005    9      58.3      58.3
      Preci_est_IDEAM IDEAM_GPCC_26075010_1.Preci_est_IDEAM
337      58.3      58.3
> IDEAM_GPCC_26075010_2[298,]
      fecha preci_GPCC secuencia ano mes preci_IDEAM Preci_est_IDEAM_1
376 01/10/2005   166.25      298 2005   10     119.7     119.7
      Preci_est_IDEAM IDEAM_GPCC_26075010_1.Preci_est_IDEAM
376     119.7     119.7
> IDEAM_GPCC_26075010_2[299,]
      fecha preci_GPCC secuencia ano mes preci_IDEAM Preci_est_IDEAM_1
415 01/11/2005   177.21      299 2005   11     118.8     118.8
      Preci_est_IDEAM IDEAM_GPCC_26075010_1.Preci_est_IDEAM
415     118.8     118.8

```

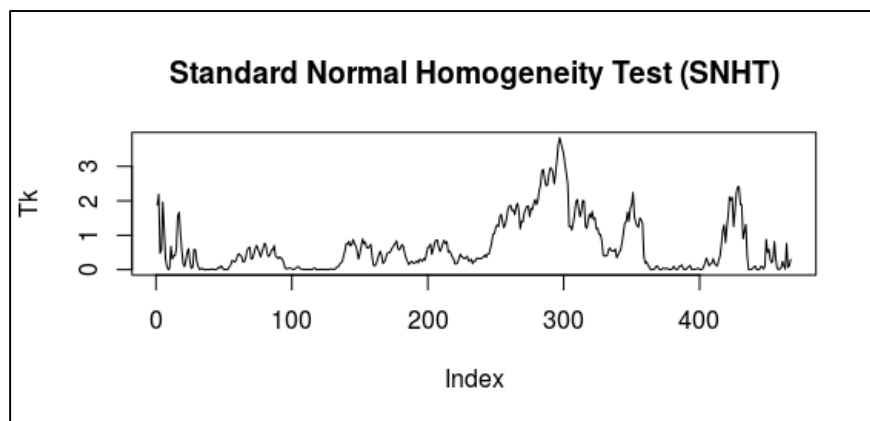


Figura 4. Ejemplo de la prueba de detección de cambios en el promedio para la serie de tiempo de preipitación mensual de ICA-Palmira

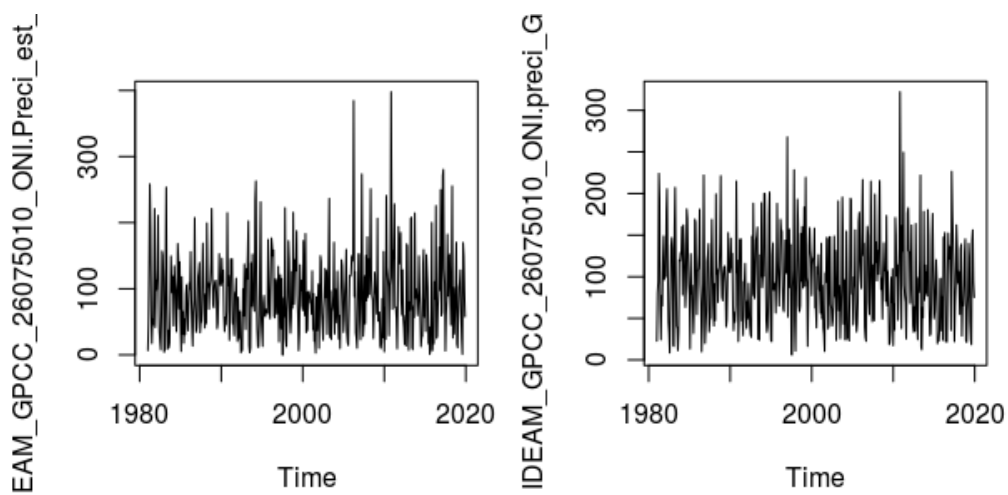


Figura 5. Serie de tiempo de precipitación mensual de ICA-Palmira con el dato faltante estimado y homogenizada (izquierda) y serie de tiempo de precipitación mensual de GPCC

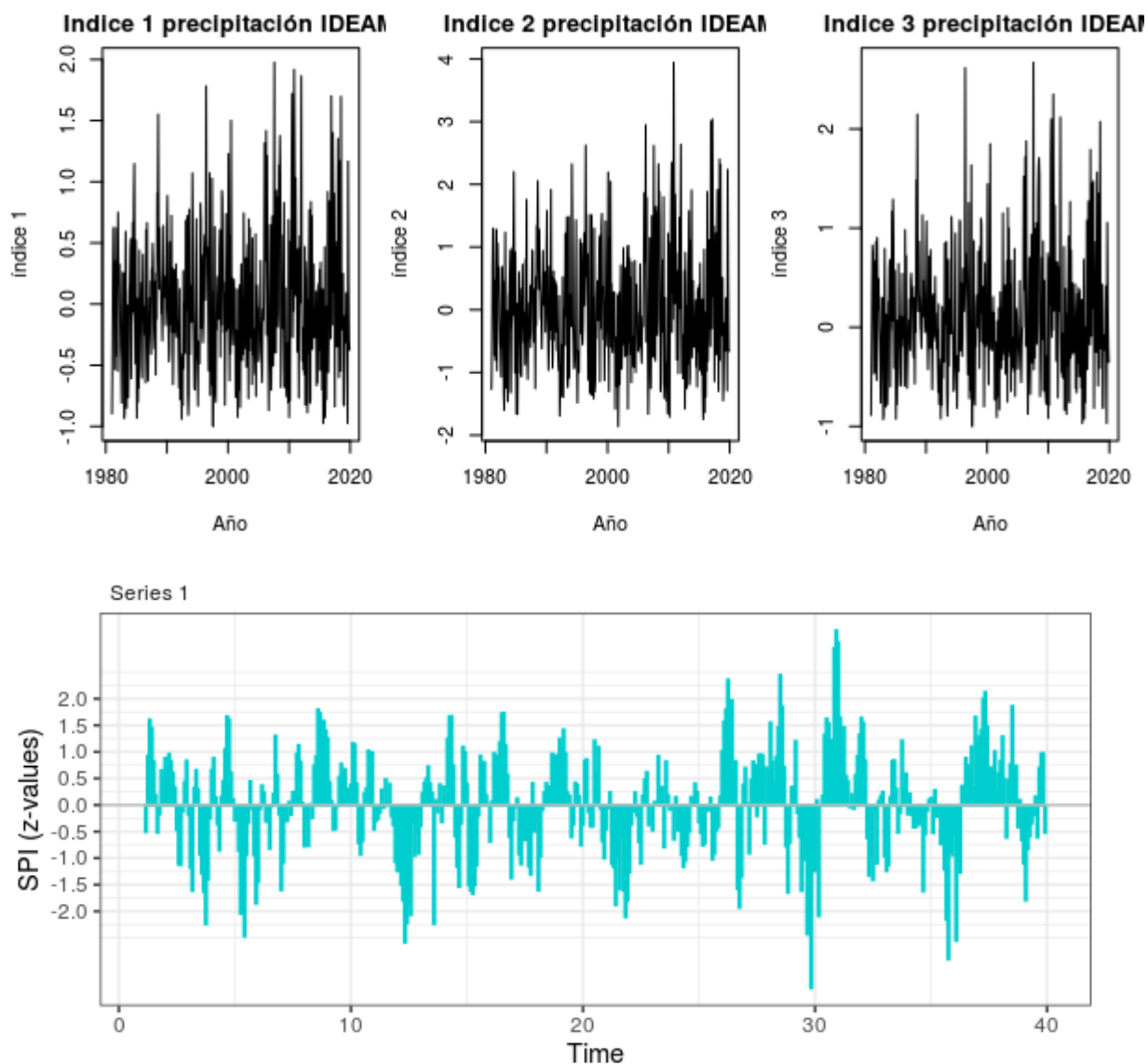


Figura 6. Serie de tiempo de índices climáticos basado en la precipitación mensual de ICA-Palmira, índice 1: anomalías respecto al promedio, índice 2: anomalías estandarizadas, índice 3: anomalías respecto a la mediana y SPI, basado en (Bernal & Pabón, 2021) y (Bernal, 2022)

Agradecimientos

Al Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), a Ruth Correa y María Inés Cubillos de la Subdirección de Meteorología. A Gustavo Cabrales por el procesamiento y descarga de los datos de GPCC, a Yenny Sánchez por la colaboración en la elaboración de algunos mapas, a Alfonso Ladino, Nicole Rivera y Fernanda Moreno, al Prof. Andreas Becker de GPCC, al Prof. Emel Vega, Diego Quintero y Prof. Ramón Giraldo de la Universidad Nacional de Colombia por las reflexiones de tipo metodológico, relacionadas con el análisis de la similitud entre series y la alternativa de un modelo espacial para estimación de datos faltantes, que se abordaron en el marco del proyecto IDEAM –UNAL en el año 2018.

Referencias

- Barbulescu, A., Maftai, C., & Bautu, E. (2010). Modeling the hydro-meteorological time series. Applications to Dobrudja region. LAMBERT Academic Publishing.
- Barrero, A., & Sabogal, N. (n.d.). Curva de dobles masas.
- Bernal, N., Correa, R. y Rangel, E. (2011). Homogenización de series de tiempo mensuales de precipitación [ponencia]. IX Congreso Colombiano de Meteorología: "200 años de la meteorología y la climatología en América Latina". Bogotá, D.C.
- Bernal, N. (2022). Notas de clase sobre análisis de variabilidad climática basado en datos de precipitación mensual.
- Bernal, N., Barrios, J., Ramos, M., Velásquez, C., Ibarra, Y., Lombana, L., Velásquez, W., & Beltrán, J. (2012, July). Propuesta Metodológica para la Homogenización de Series de Tiempo de Precipitación Mensual y su utilidad en procesos de toma de decisiones, estudio de caso Región Climatológica del Bajo Magdalena.
- Bernal, N., Barrios, J., Ramos, M., Velásquez Wendy, Beltrán, J., Ibarra, Y., & Velásquez, W. (2021). Homogenización de Series de tiempo mensuales de Precipitación y su utilidad de Estudios Climáticos y Procesos de toma de decisiones. UD.
- Bernal, N., & Pabón, D. (2021). Análisis de los Índices de precipitación utilizados para el estudio de la variabilidad climática. Universidad Del Tolima.
- Dale, D., & Hale, R. (2012). Time Series Analysis in Meteorology and Climatology: an introduction . Jhon Wiley & Sons, Ltd. .
- Eslava, J. (1994). Apuntes de Meteorología y Climatología General. Universidad Nacional de Colombia.
- Henríquez, M. (2012). Climatología ambiental de Colombia (M. P. Godoy, Ed.). Universidad Santo Tomás .
- Jaramillo, A. (2005). Clima Andino y Café en Colombia. Editorial Blanecolor Ltda.
- Lombana, L., Bernal, N., & Barrios, J. (2018). Guía para el Procesamiento de Series de tiempo de Precipitación y Temperatura: Estimación de Datos faltantes, Detección de cambios y Homogenización. UD.
- Martínez, J., Montealegre, E., & Rangel, E. (1996). Estimación de observaciones faltantes en una serie de tiempo usando modelos Arima. (IV Congreso Colombiano de Meteorología).
- Mesa, S. O. J., Poveda, J. G., & Carvajal, S. L. F. (1997). Introducción al clima de Colombia. Universidad Nacional de Colombia.
- Montealegre, E. (1990). Técnicas estadísticas aplicadas en el manejo de datos hidrológicos y meteorológicos.
- Nieto, F. H., & Ruiz, F. (2002). About a prompt strategy for estimating missing data.
- Pabón, J., & Montealegre, J. (2017). Los Fenómenos de El Niño Y de la Niña. Editorial Gente Nueva.

- Schneider, U., Hansel, S., Finger, P., Rustemeir, E., & Ziese, M. (2022). GPCC Full Data Monthly Product Version 2022 (at 0.25° at 0.5°, 1.0°, 2.5°): Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historical Data.
- Storch, H., & Zwiers, F. (1999). Statistical Analysis in Climate Research . Cambridge University Press.
- Wei S., W. (2006). Times Series Analysis: Univariate and Multivariate Methods. PEARSON EDUCATION.
- Wilks, D. (2006). Statistical Methods in the Atmospheric Sciences. Elsevier Inc.
- World Meteorological Organization (WMO). (2020). Guidelines on Homogenization (1245).