Revista de Climatología

Vol. 22 (2022): 12-23 ISSN 1578-8768

Un modelo logístico para series climatológicas en una región de Los Andes de Venezuela

Salli Villegas Rivas[1], Pedro Arturo Barboza Zelada[2], Walter Jorge Mendizabal Anticona[2], Pérez García Patricia Marlene[2], Eliana Soledad Castañeda-Núñez[2], Jamer Nórvil Mírez Toro[2], Noel Alcas Zapata[2], Jorge Luis Albarrán Gil[2], Sidanelia Flores-Silva[3], Verónica Violeta Ruiz Cutipa[4], Yary Pérez Pérez[5], José Paredes Carranza[6], Wilfredo Ruiz Camacho[7]

[1] *Universidad Nacional Experimental de los Llanos Occidentales "Ezequiel Zamora", Guanare, Portuguesa, Venezuela.*
[2] *Escuela de Posgrado, Universidad César Vallejo, Perú.*
[3] *Universidad Nacional "San Luis Gonzaga"de Ica, Perú.*
[4] *Escuela de Educación Superior Pedagógica Pública "San Francisco de Asís", Perú.*
[5] *Universidad Politécnica Territorial de Portuguesa "J J Montilla", Guanare, Portuguesa, Venezuela.*
[6] *Facultad de Tecnología Médica, Universidad Nacional de Jaén, Cajamarca, Perú.*
[7] *Facultad de Ingeniería Forestal y Ambiental, Universidad Nacional de Jaén, Cajamarca, Perú.*

Autor para correspondencia: Salli Villegas Rivas, marilovillegas@hotmail.com

(Recibido: 11-02-2022. Publicado: 20-02-2022.)

Resumen

El objetivo de esta investigación fue evaluar series de precipitación mensual mediante regresión logística multinominal con el fin de comparar la tendencia, estacionalidad y presencia de observaciones atípicas de series de precipitación mensual. Para ello se utilizaron datos de la estación meteorológica San Cristóbal en el estado Táchira y series simuladas mediante modelos de eventos extremos; Pearson tipo III, Gumbel tipo I, Log-normal y Log-Pearson tipo III. Así, para el análisis de la tendencia y estacionalidad se utilizaron gráficos de saturación de la varianza, para la detección de observaciones atípicas se utilizó la distancia de Mahalanobis (D2). Para el ajuste de modelos de eventos extremos se utilizó la estimación de máxima verosimilitud y el ajuste de densidades. De esta manera, los resultados evidenciaron una distribución asimétrica de las precipitaciones con una discontinuidad en la serie en el periodo 1973-1983, asociada a una alta variabilidad (75,75 %) como consecuencia de la presencia de observaciones atípicas causadas por errores en los registros. Así mismo, se detectaron observaciones atípicas distribuidas en la época lluviosa, asociadas al mes de agosto de 1960, junio de 1984, julio de 1985 y julio de 1989. Por otro lado, la precipitación mensual se ajustó a una distribución Pearson tipo III. La regresión logística sugirió que la única variable relacionada significativamente con la distribución teórica de la serie fue la precipitación. Así, la simulación de MonteCarlo evidenció consistencia en los estimadores de máxima verosimilitud del modelo logístico en el análisis de la precipitación mensual. Finalmente, los resultados de esta investigación mostraron que las metodologías consideradas; saturación de la varianza, distancias de Mahalanobis (D2) y regresión logística son una poderosa herramienta para el estudio de la tendencia y

homogeneidad de la precipitación mensual, detección de outliers multivariados y la comparación de series de precipitación mensual, respectivamente.

Palabras clave: *Análisis multivariado, regresión no paramétrica, precipitación mensual.*

Abstract

The objective of this paper was to evaluate a series of monthly rainfall by a multinomial logistic regression model in order to compare the trend, seasonality and presence of outliers of monthly precipitation. For which were used data from San Cristobal weather station in Tachira state and simulated series of models of extreme events; Pearson Type III, Type I Gumbel, Log-normal and log-Pearson Type III. Also, for analysis of trend and seasonal were used saturation graphics of the variance of the series, to detect outliers was used Mahalanobis distance (D2). To adjust the patterns of extreme events the maximum likelihood estimation and graphic density adjustment was used. Thus, the results showed a skewed distribution of rainfall with a discontinuity in the trend of the series in the period 1973-1983, associated with high variability (75.75 %) due to the presence of outliers caused by errors in the records. Likewise, the presence of outliers distributed mainly in the rainy season was detected, associated to August 1960, June 1984, July 1985 and July 1989. Moreover, monthly precipitation data were adjusted a Pearson type III distribution. Logistic regression suggested that the only variable significantly related to the type of theoretical distribution of the series was the precipitation. Thus, the Monte Carlo simulation showed consistency of maximum likelihood estimators of logistic model in the analysis of monthly precipitation series. Finally, the results obtained in this research showed that the methodologies considered; saturation of the variance, Mahalanobis distances (D2) and logistic regression are a powerful tool for studying the trend and homogeneity of the monthly precipitation, multivariate outlier detection and comparison of series of monthly precipitation respectively.

Keywords: *Multivariate analysis, nonparametric regression, monthly rainfall.*

1. Introducción

En Venezuela y en cualquier parte del mundo, los estudios hidrológicos son fundamentales como fuente de datos para el diseño de obras hidráulicas y para establecer áreas vulnerables ante eventos hidrometeorológicos extremos. Según Sun, et al (2006: citado por Acevedo A. y Lina A., (2009) las lluvias han sido analizadas desde hace mucho tiempo y los estudios que se han realizado han tenido diversos objetivos, sin embargo, en la mayoría de ellos el objetivo último es la determinación de los caudales máximos para el diseño de diferentes estructuras hidráulicas. La naturaleza de los eventos hidrológicos es probabilística y por tanto, para efectuar tales diseños, es necesario plantearse modelos igualmente probabilísticas que representen el comportamiento de esos eventos. En ese orden, en la realización de muchos estudios hidrológicos con fines de diseño se presentan incongruencias en la información hidrológica empleada. Razón por la cual, los modelos de la regresión logística son modelos estadísticos en los que se desea conocer la relación entre una variable dependiente cualitativa, dicotómica (regresión logística binaria o binomial) o con más de dos valores (regresión logística multinomial). En ese sentido, es importante destacar que la presente investigación se fundamenta en la búsqueda de una herramienta que de manera práctica y sencilla contribuya a mejorar el estudio de las precipitaciones, por ello, el objetivo central de esta investigación gravita en torno al hecho de evaluar series pluviométricas mediante un modelo de regresión logística multinominal con el fin de caracterizar y comparar la información relacionada con la tendencia y estacionalidad de la precipitación mensual proveniente de estaciones meteorológicas asociadas a la cuenca del Rio Torbe del estado Táchira, así como también a modelos de eventos extremos en series sintéticas.

2. Metodología

Los datos de esta investigación provienen de una serie de precipitaciones mensuales registradas en la estación San Cristóbal en el periodo 1956-2000. Para la descripción estadística de las variables registradas en la estación se tomó en consideración el tamaño de la muestra (n), años de registro, la media aritmética, la desviación estándar, la varianza, el coeficiente de variación, mínimo, cuartil (Q1), la mediana, cuartil (Q3), máxima. El análisis exploratorio de los datos (EDA) por medio gráfico se realizó con el fin de comprobar tendencias y cambios en la serie de tiempo por medio visual. Dentro del análisis exploratorio gráfico se utilizó la gráfica de serie de tiempo, el diagrama de cajas, la gráfica de doble masa y la gráfica de normalidad. Para estudiar la homogeneidad de las series de precipitación mensual se utilizó el periodo 1956 al 1991, por ser un periodo homogéneo en la serie objeto de estudio. Se realizó la detección de observaciones atípicas mediante métodos univariados, como la distancia Cook (1977), la cual mide la influencia de una observación mediante el cambio en la región elipsoidal dada en la distancia Di cuando la i-ésima observación es eliminada, así como métodos multivariados, que a menudo indican si las observaciones se encuentran relativamente lejos del centro de la distribución de los datos, específicamente la distancia de Mahalanobis, el cual es un criterio que depende de los parámetros estimados de la distribución multivariada. Se representó gráficamente las precipitaciones mensuales con el fin de ajustar modelos de eventos extremos (Log-Normal, Pearson tipo III, Log-Pearson tipo III y Gumbel tipo I) y se estimaron los parámetros para cada uno de los modelos antes mencionados mediante el método de estimación por máxima verosimilitud. Se realizó un estudio de simulación de Montecarlo con el fin de generar series de tiempo sintéticas basadas en procesos hidrológicos, los cuales son procesos estocásticos estacionarios conocidos todos los momentos de la distribución, específicamente un proceso puramente estacionario o de ruido blanco con base en tres modelos de eventos extremos (Pearson tipo III, Log-Pearson tipo III y Log-Normal). Con base en las series sintéticas de precipitación mensual generadas mediante el estudio de simulación de Montecarlo se realizó un análisis de regresión logística con el fin de construir funciones logísticas para clasificar series de precipitación mensual provenientes de distancias distribuciones teorices. Los análisis antes mencionados se realizaron con la ayuda del software R 3.3.1.

3. Resultados

En la Tabla 1 se muestra una descripción estadística de la serie de precipitaciones mensuales registradas en la estación San Cristóbal en el periodo 1956-2000, las cuales son necesarias para la estimación de los parámetros, allí se observa una precipitación promedio mensual de 133,3 mm $\pm$ 101,01 mm, con precipitaciones mínimas mensuales de 0 mm y máximas mensuales de 829,4 mm. Esta amplitud de valores de precipitación mensual se refleja en la alta variabilidad presente en toda la serie (75,75 %), resultando en características propias de un proceso estocástico, como lo son los procesos hidrológicos específicamente las series de precipitación mensual.

Tabla 1: Estadísticas para precipitaciones mensuales registradas en la estación meteorológica San Cristóbal en el periodo 1956-2000.

Estadística	Valor
N	552
Mínimo (mm)	0
Máximo (mm)	829,4
Primer cuartil Q1 (mm)	53,7
Mediana (mm)	116,5
Tercer cuartil Q3 (mm)	196,5
Media (mm)	133,3
Varianza (mm2)	10202,28
Desviación estándar (mm)	101,01
Coeficiente de variación (%)	75,75

En la Figura 1 se muestran la distribución de las precipitaciones mensuales registradas en la estación meteorológica San Cristóbal en el periodo 1956-2000, mediante un histograma y el ajuste de una función de densidad para dicha serie, estos evidencian una tendencia de los datos de las precipitaciones a distribuirse de forma asimétrica, lo cual coincide con lo reportado en la literatura en relación al tipo de distribución de las precipitaciones.

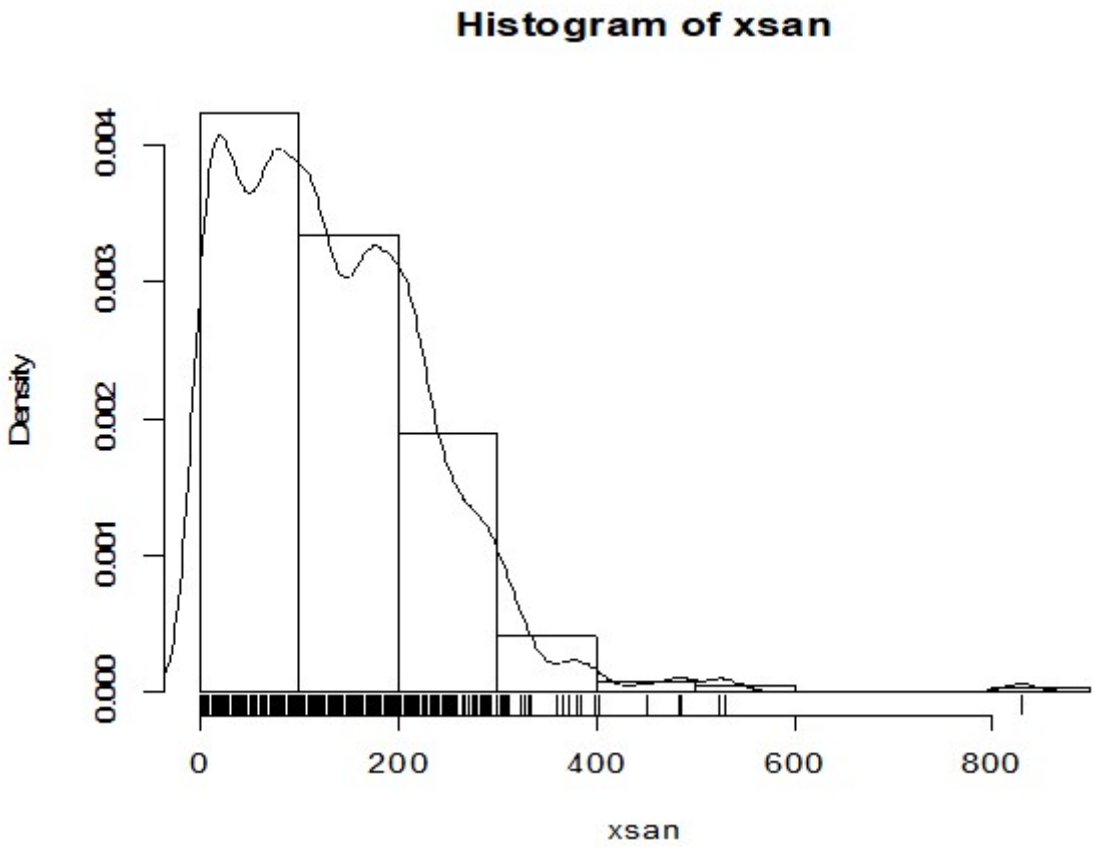

Figura 1: Distribución de las precipitaciones mensuales registradas en la estación meteorológica San Cristóbal en el periodo 1956-2000.

En la Figura 2 se muestra un gráfico de la saturación de la varianza de precipitaciones mensuales registradas en la estación San Cristóbal en el periodo 1951-2000, allí se observa que la varianza

correspondiente a la serie de precipitación mensual estudiada comienza descender a partir del año 1960 mostrando una tendencia a estabilizarse en el periodo 1960-2000, por lo que el análisis posterior debe limitarse a dicho periodo.

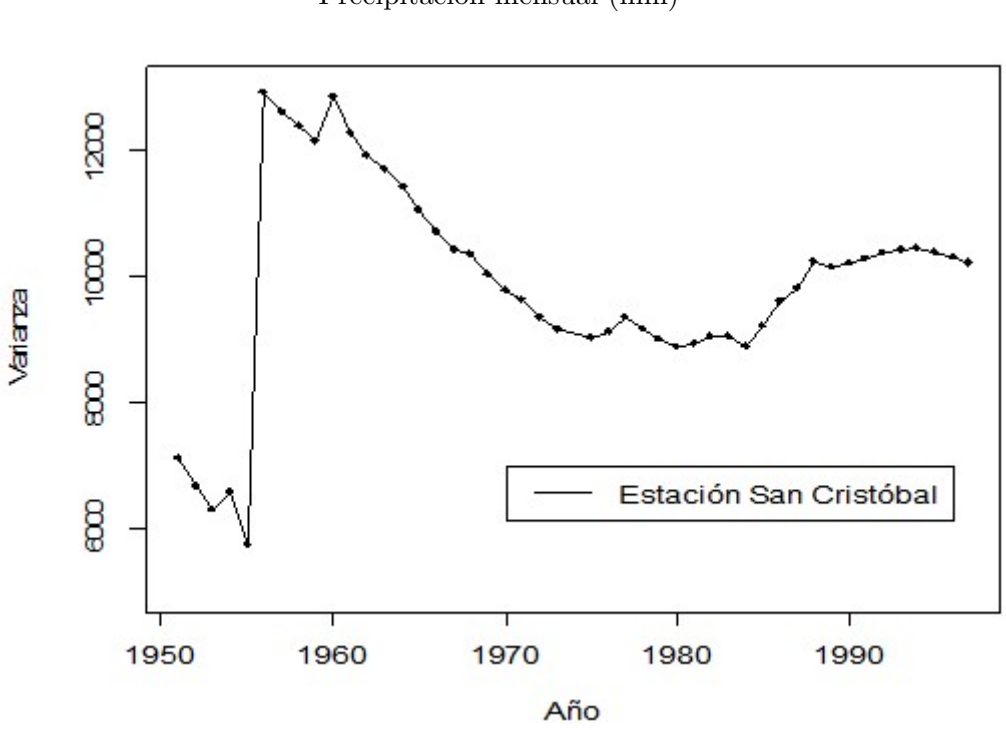

Figura 2: Saturación de la varianza de precipitaciones mensuales registradas en la estación San Cristóbal en el periodo de 1951-2000.

En la Figura 3 se muestra las precipitaciones mensuales registradas en la estación San Cristóbal en el periodo de 1951-2000, donde se observa una discontinuidad en la tendencia de la serie en el periodo 1973-1983, lo que incide en la alta variabilidad de las precipitaciones mensuales (75,75 %), con el subsecuente efecto observado en la saturación de la varianza de las mismas. Este comportamiento en la tendencia de la serie para ese período está asociado a la presencia de observaciones atípicas (outliers), causadas por errores en los registros, para lo cual se recomienda además de la detección de outlier mediante métodos multivariados, y el ajuste y estimación de parámetros mediante metodologías que consideren la presencia de verosimilitud irregulares propias de los procesos hidrológicos y distribuciones probabilísticas asimétricas, como es el caso el caso de los modelos de eventos extremos (Log-Normal, Pearson tipo III, Log-Pearson tipo III, Gumbel tipo I), también se sugiere cortar la serie y eliminar el período que presenta la discontinuidad en la tendencia (1973-1983) con el fin de reconstruir la serie de precipitaciones mensuales en todo el periodo.

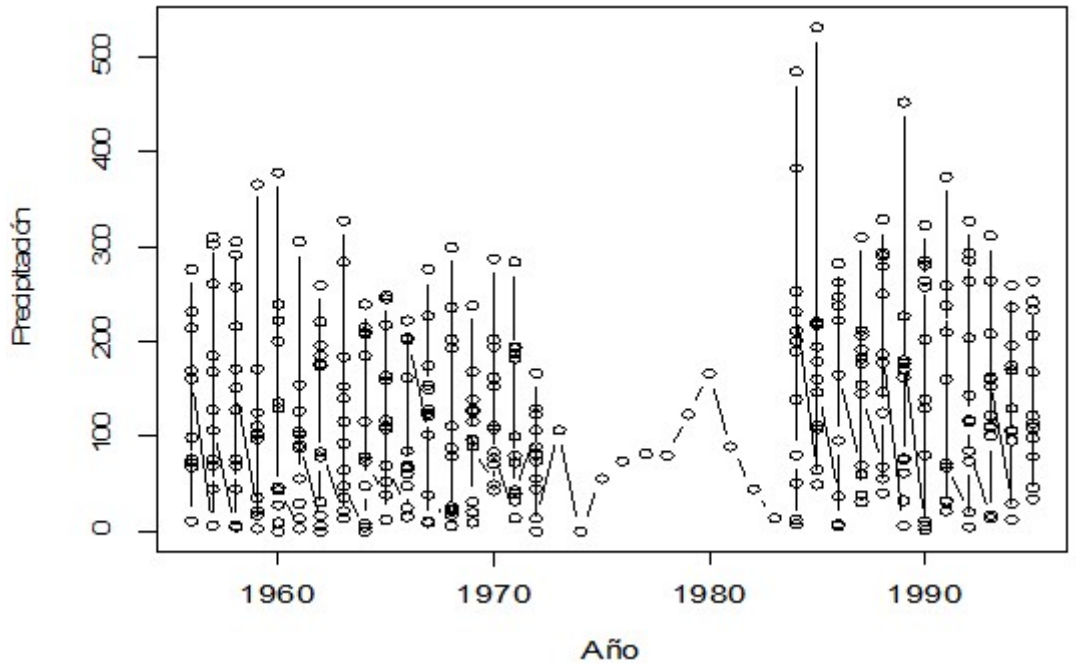

Figura 3: Precipitaciones mensuales registradas en la estación San Cristóbal en el periodo de 1951-2000.

En la Figura 4. Se muestran los resultados de la detección multivariada de outlier mediante la Distancia de Mahalanobis (D2) vs Quantiles de la distribución Chi- Cuadrado registradas en la estación San Cristóbal en el periodo de 1956-2000, allí se observan que existen cuatro observaciones que pudieran ser consideradas observaciones atípicas, dado que se alejan considerablemente del centro de masa (centroide o media multivariada), las cuales se describen con detalle en la Tabla 2, en donde se observa que estas observaciones atípicas están asociadas a distancias de Mahalanobis relativamente grandes (D2 ¿10), y se distribuyen en la época de lluvia, con precipitaciones elevadas ocurridas en agosto del año 1960 (378,7 mm), junio de 1984 (484 mm), julio de 1985 (531 mm) y julio de 1989 (451,8 mm).

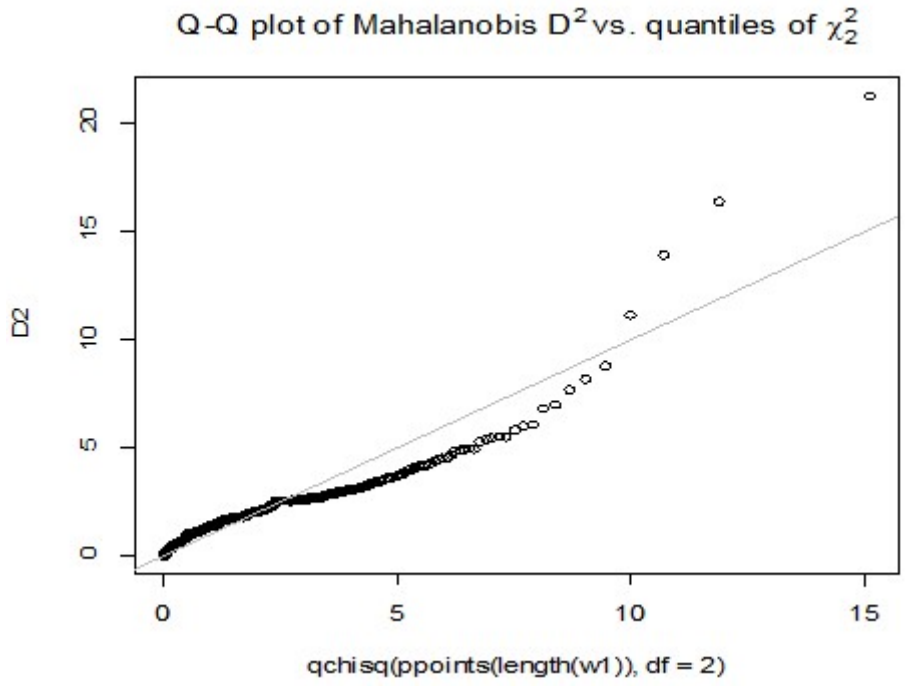

Figura 4: Distancia de Mahalanobis vs Quantiles de la distribución Chi- Cuadrado en series de precipitaciones mensuales registradas en la estación San Cristóbal en el periodo de 1951-2000.

Tabla 2: Descripción de observaciones atípicas en series de precipitaciones mensuales registradas en la estación San Cristóbal en el periodo de 1951-2000.

Año	**Mes**	**Precipitación (mm)**	**Distancia de Mahalanobis (D^2)**
1960	Agosto	378,7	11,4
1984	Junio	484	16,39
1985	Julio	531	21,29
1989	Julio	451,8	13,91

En la Tabla 3 se muestran los resultados del ajuste y estimación de parámetros y ajuste de modelos de eventos extremos en series de precipitación mensual de la estación meteorológica San Cristóbal en el período 1956-2000, allí se observa que el test de Kolmogorov- Smirnoff sugiere que el modelo que muestra el mejor ajuste al conjunto de datos de precipitación mensual es el Pearson tipo III, estos resultados son verificados al observar la Figura 5, donde se muestran las densidades para los cuatro modelos de eventos extremos, allí se observa que el modelo Pearson tipo III es el que mejor se ajusta al histograma de precipitaciones mensuales de la estación San Cristóbal. Estos resultados verifican lo señalado por algunos autores, quienes afirman que el modelo Pearson tipo III o Gamma de tres parámetros es el que mejor se ajusta a la distribución de las precipitaciones mensuales.

Tabla 3: Ajuste y estimación por máxima verosimilitud de parámetros de modelos de eventos extremos en series de precipitación mensual de la estación meteorológica San Cristóbal en el período 1956-2000.

Modelo	**Parámetro estimado**	Bondad de ajuste (Test de Kolmogorov-Smirnoff)	
		Estadístico de prueba (D)	**Significación (P Valor)**
Log-Normal	$\mu = 4,459281$ $\sigma = 1,275493$	0,15533	1,019e-7
Pearson tipo III	$\alpha = 1,230967$ $x_0 = 0,0090068$	0,089845	0,007263
Log-Pearson tipo III	$\alpha = 10,75065$ $x_0 = 2,339255$	0,19097	1,894e-11
Gumbel tipo I	$\alpha = 91,01674$ $\beta = 4,360523$	0,53869	2,2e-16

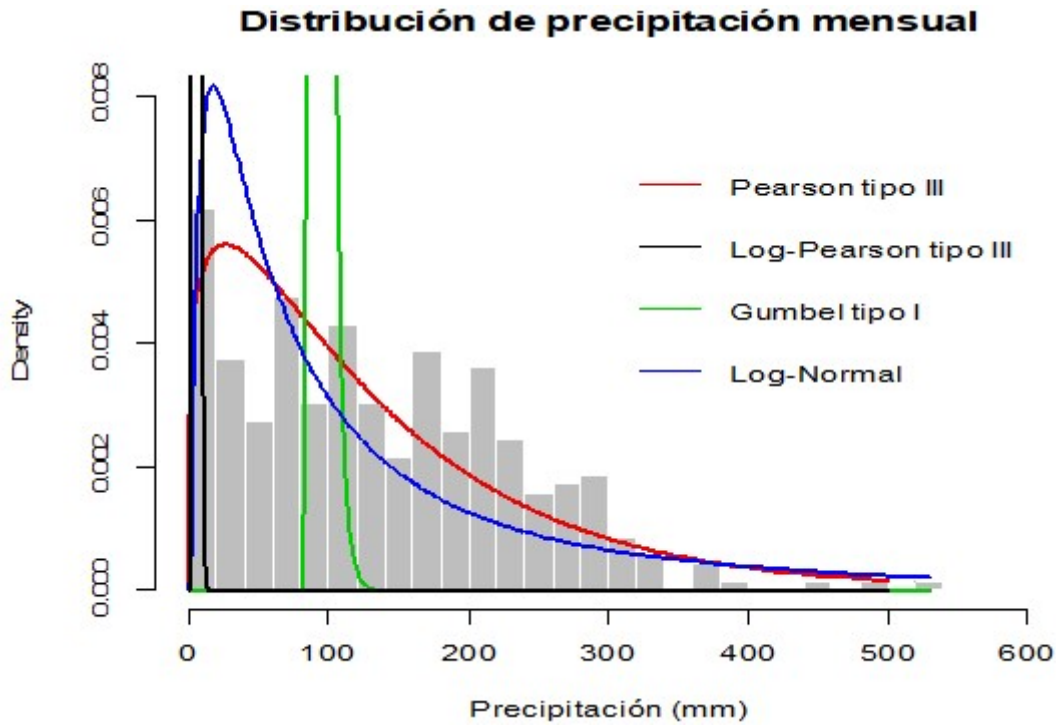

Figura 5: Ajuste de modelos de eventos extremos en una serie de precipitaciones mensuales registradas en la estación meteorológica San Cristóbal en el periodo 1951-2000.

En la Tabla 4 se muestran los resultados de la regresión logística sobre tres series de precipitaciones mensuales simuladas provenientes de una distribución Pearson tipo III, Log-Pearson y Gumbel con precipitación, temperatura y humedad como regresoras como las que se muestran en la Figura 6, allí se observa que el estadístico Z de Wald para el análisis individual de las variables regresoras sugiere que la única variable relacionada significativamente con el tipo de distribución teórica de la serie es la precipitación.

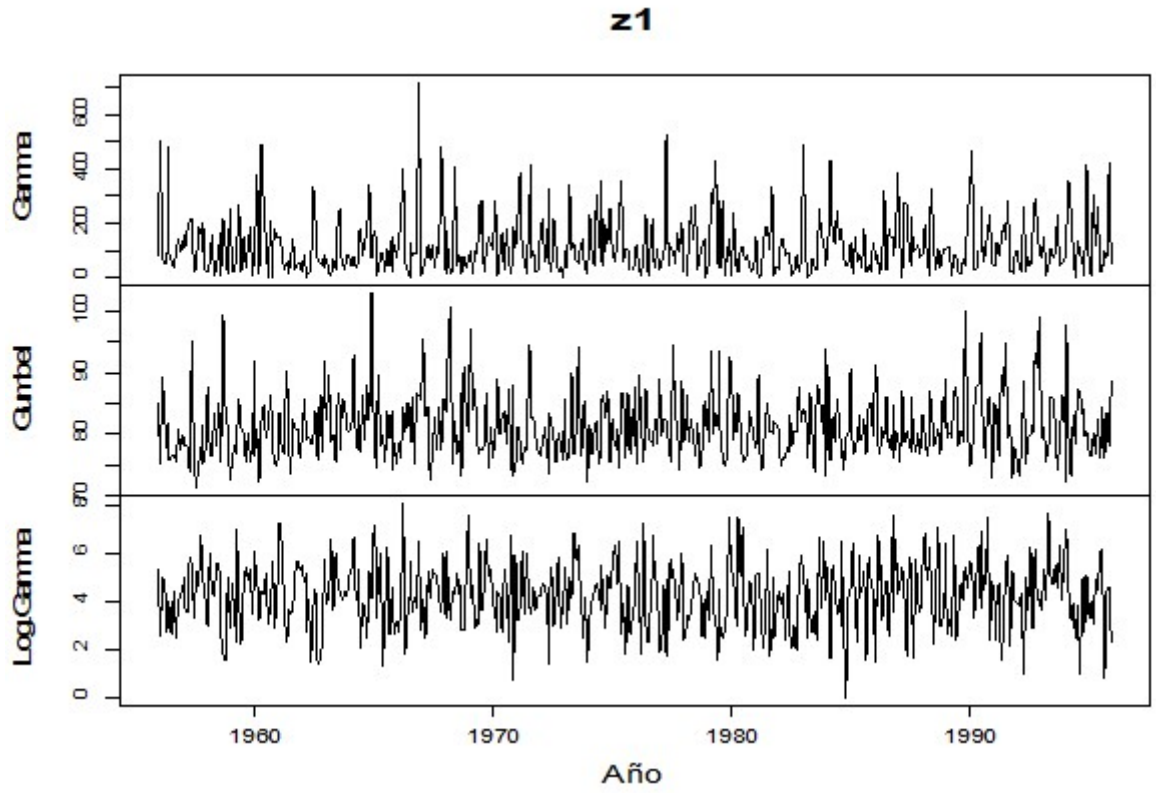

Figura 6: Series de precipitaciones mensuales reconstruida (simulada mediante una distribución Gamma o Pearson tipo III, Log-Pearson y Gumbel) en la estación meteorológica San Cristóbal en el periodo 1956-2000.

Tabla 4: Regresión logística con tres series de precipitaciones mensuales simuladas provenientes de una distribución Pearson tipo III, Log-Pearson y Gumbel con precipitación, temperatura y humedad como regresoras.

Función logística 1 Log-normal / Pearson III							
	Coeficiente	**Desv. Estándar coeficientes**	**Z**	**P**	**Odds Ratio**	**IC 95 % Límite inferior**	**IC 95 % Límite superior**
Intercepto	-0,61638	4,70181	-0,13	0,896			
Precipitación	-0,596177	0,08172	-7,30	0,000	0,55	0,47	0,65
Temperatura	0,107136	0,09822	1,09	0,275	1,11	0,92	1,35
Humedad	0,056729	0,04637	1,22	0,221	1,06	0,97	1,16
Función logística 2 Gumbel I / Pearson III							
Intercepto	0,889606	1,7004	0,52	0,601			
Precipitación	-0,007238	0,001172	-6,18	0,000	0,99	0,99	1,00
Temperatura	-0,000882	0,035366	-0,02	0,980	1,00	0,93	1,07
Humedad	-0,002259	0,017184	-0,13	0,895	1,00	0,96	1,03

El modelo sería:

$$P(\text{Log} - \text{normal/precipitación (mm)}) = \frac{(e^{6,53351 - 0,585475 * \text{precipitación}})}{(1 + e^{6,53351 - 0,585475 * \text{precipitación}})}$$

$$\text{P(Gumbel I/precipitación (mm))} = \frac{(e^{0,681652 - 0,0072367 * \text{precipitación}})}{(1 + e^{0,681652 - 0,0072367 * \text{precipitación}})}$$

En la Tabla 5 se muestran los resultados de las pruebas de bondad de ajuste de un modelo de regresión logística con tres series de precipitaciones mensuales simuladas provenientes de una distribución Pearson tipo III, Log-Pearson y Gumbel con precipitación, temperatura y humedad como regresoras, allí se observa que los resultados de los estadísticos asociados a la razón de verosimilitud mejoran conforme se incrementa el tamaño de la muestra, lo que evidencia la consistencia de los estimadores de máxima verosimilitud del modelo logístico en el análisis de series de precipitación mensual.

Tabla 5: Bondad de ajuste de un modelo de regresión logística con tres series de precipitaciones mensuales simuladas provenientes de una distribución Pearson tipo III, Log-Pearson y Gumbel con precipitación, temperatura y humedad como regresoras.

Muestral (n)	Log-verosimilitud	G	P valor
10	-12,806	40,306	0,000
30	-38,558	120,634	0,000
50	-71.89	186,405	0,000
100	-164,551	330,065	0,000
200	-317,966	682,403	0,000
300	-473,993	1029,517	0,000
400	-636,465	1363,739	0,000
480	-761,46	1641,083	0,000

5. Conclusiones

La regresión logística sobre tres series de precipitaciones mensuales simuladas provenientes de una distribución Pearson tipo III, Log-Pearson y Gumbel con precipitación, temperatura y humedad como regresoras sugirió que la única variable relacionada significativamente con el tipo de distribución teórica de la serie fue la precipitación. Los resultados de la regresión logística con tres series de precipitaciones mensuales simuladas provenientes de una distribución Pearson tipo III, Log-Pearson y Gumbel con precipitación, temperatura y humedad como regresoras, mostraron como los estadísticos asociados a la razón de verosimilitud mejoraron conforme se incrementó el tamaño de la muestra, lo que evidenció la consistencia de los estimadores de máxima verosimilitud del modelo logístico en el análisis de series de precipitación mensual. Finalmente, en virtud de los resultados obtenidos en esta investigación se recomienda considerar las metodologías presentadas, como es el caso de la saturación de la varianza como alternativa para el estudio de la tendencia y homogeneidad en series de precipitación mensual, así como el uso de las distancias de Mahalanobis (D2) para la detección de outliers multivariados y la regresión logística como una poderosa herramienta para la comparación de series de precipitación mensual.

Referencias bibliográficas

Acevedo A, Lina A (2009): Estimación hidrológica bajo escenarios de cambio climático en Colombia. Tesis de maestría en Aprovechamiento de Recursos Hidráulicos. Universidad Nacional de Colombia, Facultad de Minas. Medellin, Colombia.

Alexanderson H (1986): A Homogeneity Test to Precipitation Data, *Journal of Climatology*, 6:661–675.

Beckman R, Cook RD (1983): Outlier... (with Discussion), Technometics2b, 119-149.

Birkes D, Dodge Y (1993): Alternative methods of regression. New York: John Wiley and Sons. 240 p.

Martelo MT (2004): Consecuencias ambientales generales del cambio climático en Venezuela, Trabajo de ascenso, Universidad Central de Venezuela, Facultad de Agronomía, Maracay, Venezuela.

Sánchez J (1999): Manual de análisis estadístico de los datos. Segunda edición. Alianza Editorial S.A. Madrid.

Santiago de la Fuente Fernández (2011): Regresión Logística. Facultad de Ciencias Económicas y Empresariales. UAM. Madrid. 29 p.

Searcy JH, Hardison CH. Manual of hidrology 1. General surface-water techniques. Double-mass curves. Washington (Estados Unidos). Department of Agriculture, 1963. P39-40.

Smith R, Campuzano C (2000): Análisis exploratorio para la detección de cambios y tendencias en series hidrológicas. XIV Seminario Nacional de Hidráulica e Hidrología.

Wendy AB (1996): Introduction to logistic regression models with worked forestry examples biometrics Information. Handbook No 7. Province of British Columbia. Ministry of Forests Research Program. 147 p.