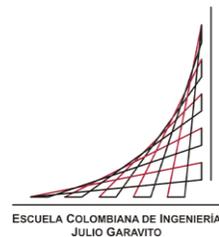


Maestría en Ingeniería Civil

Análisis comparativo de distribuciones estadísticas para el cálculo de una variable hidrológica

Wilmar Hernán Barbosa Hernández

Bogotá D.C., 28 de mayo de 2020



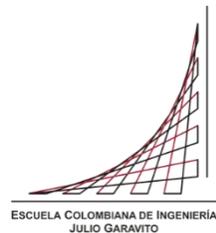
**Análisis comparativo de distribuciones estadísticas para el
cálculo de una variable hidrológica**

**Tesis para optar al título de magister en Ingeniería Civil con
énfasis en recursos hidráulicos y medio ambiente**

Héctor Alfonso Rodríguez Díaz

Director

Bogotá D.C., 28 de mayo de 2019



La tesis de maestría titulada “Análisis comparativo de distribuciones estadísticas para el cálculo de variables hidrológicas” presentada por Wilmar Hernán Barbosa Hernández, cumple con los requisitos establecidos para optar al título de Magister en Ingeniería Civil con énfasis en recursos hidráulicos y medio ambiente.

Director de la tesis

Ing. Héctor Alfonso Rodríguez Díaz

Jurado

Ing. Mónica Andrea Vargas Solla

Jurado

Ing. Andrés Humberto Otálora Carmona

Bogotá D.C., 28 de abril de 2019

Dedicatoria

Dedico este trabajo de grado a José y María, mis padres, por su apoyo incondicional para alcanzar las metas que me he propuesto en la vida.

Agradecimientos

A la Escuela Colombiana de Ingeniería, y sus profesores que aportaron con su conocimiento y experiencia en mi formación profesional.

A mi director de tesis, ing. Alfonso Rodríguez por su amplio conocimiento, quien brindó las directrices requeridas para finalizar el presente trabajo de grado.

A la codirectora de tesis, Ing. Mónica Vargas por su gestión incondicional, y aportes durante el desarrollo del trabajo de grado.

TABLA DE CONTENIDO

1. INTRODUCCIÓN.....	15
1.1. Contexto	15
1.2. Objetivos.....	16
2. MARCO TEÓRICO Y ESTADO DEL ARTE	17
2.1. Marco Teórico.....	17
2.1.1. Caracterización de series de datos máximos	17
2.1.2. Datos faltantes.....	18
2.1.2.1. Datos completados por regresión lineal.....	18
2.1.2.2. Datos completados por razones de distancias	19
2.1.2.3. Datos completados por promedios vecinales	20
2.1.2.4. Método de las proporciones.....	21
2.1.2.5. Método de correlación lineal múltiple.....	21
2.1.3. Distribución empírica de datos.....	22
2.1.4. Análisis de series de tiempo de variables hidroclimatológicas.....	23
2.1.4.1. Prueba F	24
2.1.4.2. Prueba T	25
2.1.4.3. Prueba de distribución libre CUSUM	25
2.1.4.4. Prueba Rank Sum.....	26
2.1.4.5. Prueba de regresión lineal	27
2.1.4.6. Prueba Mann – Kendall	28
2.1.4.7. Prueba de Autocorrelación	29
2.1.5. Distribuciones de uso común en hidrología	30
2.1.5.1. Distribución normal	30
2.1.5.2. Distribución log normal	32
2.1.5.3. Distribución Exponencial.....	32
2.1.5.4. Distribución Gamma	33
2.1.5.5. Distribución Pearson Tipo III.....	35

2.1.5.6.	Distribución Log Pearson Tipo III	36
2.1.5.7.	Distribución de valor extremo.....	36
2.1.6.	Distribuciones estadísticas inusuales aplicadas en hidrología.....	38
2.1.7.	Métodos de ajuste a una distribución de probabilidad	52
2.1.7.1.	Método de los momentos.....	52
2.1.7.2.	Método de la máxima verosimilitud.....	53
2.1.8.	Pruebas de bondad de ajuste	53
2.1.8.1.	Prueba de Kolmogorov – Smirnov	54
2.1.8.2.	Prueba de Anderson-Darling	54
2.1.8.3.	Prueba de chi Cuadrado	55
2.2.	Estado del arte	56
3.	METODOLOGÍA.....	60
3.1.	Programas utilizados en el trabajo de grado	60
3.1.1.	Trend 1.0.2	60
3.1.2.	EasyFit 5.6.....	60
3.2.	Información hidroclimatológica preliminar.....	61
3.3.	Procedimiento para seleccionar las estaciones hidroclimatológicas	62
3.3.1.	Ejemplo para pruebas paramétricas	65
3.3.2.	Ejemplo para pruebas no paramétricas	70
3.4.	Variables hidroclimatológicas seleccionadas.....	77
3.5.	Estaciones empleadas y caracterización del área de estudio	78
3.6.	Procedimiento para aplicar las distribuciones estadísticas y seleccionar los mejores ajustes	84
4.	RESULTADOS AL APLICAR LAS DISTRIBUCIONES ESTADÍSTICAS.....	90
4.1.	Precipitación total anual con prueba de Kolmogorov-Smirnov	90
4.2.	Precipitación total anual con prueba de Anderson – Darling	101
4.3.	Precipitación máxima en 24 horas con prueba de Kolmogorov-Smirnov.....	112
4.4.	Precipitación máxima en 24 horas con prueba de Anderson – Darling	123
5.	CONCLUSIONES Y RECOMENDACIONES	134

6. REFERENCIAS BIBLIOGRÁFICAS	137
-------------------------------------	-----

INDICE DE TABLAS

Tabla 1 Ecuaciones de probabilidad de excedencia empíricas.....	22
Tabla 2 Funciones de densidad para las distribuciones de valor extremo tipo I (Gumbel), tipo II (Fréchet) y tipo III (Weibull)	37
Tabla 3 Distribuciones estadísticas aplicadas en el presente trabajo de grado	43
Tabla 4 Estudios previos sobre distribuciones estadísticas basados en análisis clásicos y análisis regional de frecuencias	58
Tabla 5 Datos originales precipitaciones totales mensuales, estación Núñez (21190330)	65
Tabla 6 Datos completados de precipitaciones totales mensuales, estación Núñez (21190330)	66
<i>Tabla 7 Prueba F con significancia del 5% para la serie de precipitaciones totales anuales, estación Núñez (21190330).....</i>	<i>68</i>
Tabla 8 Prueba t con significancia del 5% para la serie de precipitaciones totales anuales, estación Núñez (21190330).....	68
Tabla 9 Regresión lineal con significancia del 5% para la serie de precipitaciones totales anuales, estación Núñez (21190330)	69
Tabla 10 Datos originales precipitaciones máximas en 24 horas, estación Villapinzón Granja (21201640).....	71
Tabla 11 Datos completados precipitaciones máximas en 24 horas, estación Villapinzón Granja (21201640).....	72
Tabla 12 Estaciones descartadas, con altitudes y/o precipitaciones totales por fuera de 1.5 veces la desviación estándar de la muestra para precipitación total anual	79
Tabla 13 Estaciones descartadas, con altitudes y/o precipitaciones totales por fuera de 1.5 veces la desviación estándar de la muestra para precipitación máxima en 24 horas	80
Tabla 14 Estaciones hidroclimatológicas definitivas para precipitación total anual, empleadas en el presente trabajo de grado	81
Tabla 15 Estaciones hidroclimatológicas definitivas para precipitación máxima en 24 horas, empleadas en el presente trabajo de grado	83
Tabla 16 Puntajes obtenidos por las distribuciones estadísticas (61), organizados para los 5 mejores ajustes, para cada una de las estaciones (20) seleccionadas. Series anuales de precipitación total. Prueba de Kolmogorov Smirnov.....	90
Tabla 17 Puntajes totales obtenidos por las distribuciones estadísticas (61). Series anuales de precipitación total. Prueba de Kolmogorov Smirnov.....	93
Tabla 18 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Kolmogorov-Smirnov	96
Tabla 19 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Kolmogorov-Smirnov.....	97

Tabla 20 Puntajes obtenidos por las distribuciones estadísticas (61), organizados para los 5 mejores ajustes, para cada una de las estaciones (20) seleccionadas. Series anuales de precipitación total. Prueba de Anderson-Darling.....	101
Tabla 21 Puntajes totales obtenidos por las distribuciones estadísticas (61). Series anuales de precipitación total. Prueba de Anderson-Darling	104
Tabla 22 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Anderson-Darling.....	107
Tabla 23 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Anderson-Darling	108
Tabla 24 Puntajes obtenidos por las distribuciones estadísticas (61), organizados para los 5 mejores ajustes, para cada una de las estaciones (20) seleccionadas. Series anuales de precipitación máxima en 24 horas. Prueba de Kolmogorov-Smirnov.....	112
Tabla 25 Puntajes totales obtenidos por las distribuciones estadísticas (61). Series anuales de precipitación máxima en 24 horas. Prueba de Kolmogorov-Smirnov	115
Tabla 26 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Kolmogorov-Smirnov	118
<i>Tabla 27 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Kolmogorov-Smirnov.....</i>	<i>119</i>
<i>Tabla 28 Puntajes obtenidos por las distribuciones estadísticas (61), organizados para los 5 mejores ajustes, para cada una de las estaciones (20) seleccionadas. Series anuales de precipitación máxima en 24 horas. Prueba de Anderson-Darling.....</i>	<i>123</i>
Tabla 29 Puntajes totales obtenidos por las distribuciones estadísticas (61). Series anuales de precipitación máxima en 24 horas. Prueba de Anderson-Darling.....	126
Tabla 30 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Anderson-Darling.....	129
Tabla 31 <i>Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Anderson-Darling.....</i>	<i>130</i>

INDICE DE FIGURAS

Figura 1 Disposición espacial para completar datos, por razones de distancia. Fuente: Pizarro et al, 1993.	19
Figura 2 Disposición espacial para completar datos, por promedios vecinales. Fuente: Pizarro et al, 1993.	20
Figura 3 Función de densidad de la distribución normal estándar	31
Figura 4 Función de densidad de la distribución exponencial	33
Figura 5 Función de densidad Gamma para diferentes parámetros	34
Figura 6 Estaciones hidroclimatológicas en Cundinamarca	61
Figura 7 Categoría para las 624 estaciones activas de Cundinamarca en régimen hidrológico Andino	62
Figura 8 Algoritmo para establecer o descartar una estación hidroclimatológica - fuente: Elaboración propia	64
Figura 9 Prueba de normalidad en Excel con 6 intervalos mediante Kolmogorov-Smirnov, para la serie de precipitaciones totales anuales, estación Núñez (21190330).....	67
Figura 10 Prueba autocorrelación con software Trend v1.0.2. para la serie de precipitaciones totales anuales, estación Núñez (21190330).....	70
Figura 11 Prueba de normalidad en Excel, para la serie anual de precipitaciones máximas en 24 horas, estación Villapinzón Granja (21201640).....	73
Figura 12 Prueba CuSum con software Trend v1.0.2 para la serie anual de precipitaciones máximas en 24 horas, estación VillaPinzón Granja (21201640)	74
Figura 13 Prueba RankSum con software Trend v.1.0.2 para la serie anual de precipitaciones máximas en 24 horas, estación VillaPinzón Granja (21201640)	75
Figura 14 Prueba Mann-Kendall con software Trend v.1.0.2 para la serie anual de precipitaciones máximas en 24 horas, estación VillaPinzón Granja (21201640)	76
Figura 15 Prueba autocorrelación con software Trend v.1.0.2 para la serie de precipitaciones totales anuales, estación Núñez (21190330).....	77
Figura 16 Estaciones seleccionadas en Cundinamarca para estudiar la variable de precipitación total anual – Fuente: Elaboración propia en ArcGIS	82
Figura 17 Estaciones seleccionadas en Cundinamarca para estudiar la variable de precipitación total anual – Fuente: Elaboración propia en ArcGIS	84
Figura 18 Interfaz gráfica del programa de distribuciones estadísticas EASYFIT 5.6.....	85
Figura 19 Serie de precipitación total anual estación Núñez ingresada en el software EasyFit (21190330).....	86
Figura 20 Distribuciones estadísticas ordenadas de mejor a peor ajuste por Kolmogorov-Smirnov para la estación Núñez (21190330)	86

Figura 21 Distribuciones estadísticas ordenadas de mejor a peor ajuste por Anderson - Darling para la estación Núñez (21190330).....	87
Figura 22 Valor de los parámetros de las distribuciones aplicadas a la serie de precipitaciones totales anuales de la estación Núñez (21190330)	88
Figura 23 Valores de los estadísticos por cada distribución estadística y posición en el ranking ...	89
<i>Figura 24</i> Puntajes totales obtenidos por las distribuciones estadísticas (61) para las series de precipitaciones totales anuales. Prueba de Kolmogorov Smirnov <i>Figura 25</i> Número de apariciones en el ranking por las distribuciones estadísticas (61) para las series de precipitaciones totales anuales. Prueba de Kolmogorov Smirnov	95
Figura 26 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Kolmogorov-Smirnov. Distribución de referencia Normal.....	96
Figura 27 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Kolmogorov-Smirnov. Distribución de referencia normal	97
Figura 28 Distribuciones que aparecen en primera posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación total anual (20)	98
Figura 29 Distribuciones que aparecen en segunda posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación total anual (20)	98
Figura 30 Distribuciones que aparecen en tercera posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación total anual (20)	99
Figura 31 Distribuciones que aparecen en cuarta posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación total anual (20)	99
Figura 32 Distribuciones que aparecen en quinta posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación total anual (20)	100
Figura 33 Puntajes totales obtenidos por las distribuciones estadísticas (61) para las series de precipitaciones totales anuales. Prueba de Anderson-Darling	106
<i>Figura 34</i> Número de apariciones en el ranking por las distribuciones estadísticas (61) para las series de precipitaciones totales anuales. Prueba de Anderson-Darling	106
Figura 35 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Anderson-Darling. Distribución de referencia Normal.	107
Figura 36 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Anderson-Darling. Distribución de referencia normal	108
Figura 37 Distribuciones que aparecen en primera posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación total anual (20).....	109

Figura 38 Distribuciones que aparecen en segunda posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación total anual (20).....	109
Figura 39 Distribuciones que aparecen en tercera posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación total anual (20).....	110
Figura 40 Distribuciones que aparecen en cuarta posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación total anual (20).....	110
Figura 41 Distribuciones que aparecen en quinta posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación total anual (20).....	111
<i>Figura 42 Puntajes totales obtenidos por las distribuciones estadísticas (61) para las series de precipitación máxima en 24 horas. Prueba de Kolmogorov-Smirnov.....</i>	<i>117</i>
<i>Figura 43 Número de apariciones en el ranking por las distribuciones estadísticas (61) para las series de precipitación máxima en 24 horas. Prueba de Kolmogorov-Smirnov.....</i>	<i>117</i>
<i>Figura 44 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Kolmogorov-Smirnov. Distribuciones Gumbel y Log Pearson 3 de referencia</i>	<i>118</i>
<i>Figura 45 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Kolmogorov-Smirnov. Distribuciones Gumbel y Log Pearson 3 de referencia.....</i>	<i>119</i>
Figura 46 Distribuciones que aparecen en primera posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación máxima en 24 horas (20).	120
Figura 47 Distribuciones que aparecen en segunda posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación máxima en 24 horas (20).	120
Figura 48 Distribuciones que aparecen en tercera posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación máxima en 24 horas (20).	121
Figura 49 Distribuciones que aparecen en cuarta posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación máxima en 24 horas (20).	121
Figura 50 Distribuciones que aparecen en quinta posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación máxima en 24 horas (20).	122
<i>Figura 51 Puntajes totales obtenidos por las distribuciones estadísticas (61) para las series de precipitación máxima en 24 horas. Prueba de Anderson-Darling.....</i>	<i>128</i>
<i>Figura 52 Número de apariciones en el ranking por las distribuciones estadísticas (61) para las series de precipitación máxima en 24 horas. Prueba de Anderson-Darling.....</i>	<i>128</i>
<i>Figura 53 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Anderson-Darling. Distribuciones Gumbel y Log Pearson 3 de referencia</i>	<i>129</i>
<i>Figura 54 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Anderson-Darling. Distribuciones Gumbel y Log Pearson 3 de referencia.....</i>	<i>130</i>

Figura 55 Distribuciones que aparecen en primera posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación máxima en 24 horas (20).....	131
Figura 56 Distribuciones que aparecen en segunda posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación máxima en 24 horas (20).....	131
Figura 57 Distribuciones que aparecen en tercera posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación máxima en 24 horas (20).....	132
Figura 58 Distribuciones que aparecen en cuarta posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación máxima en 24 horas (20).....	132
Figura 59 Distribuciones que aparecen en quinta posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación máxima en 24 horas (20).....	133

1. INTRODUCCIÓN

En el capítulo 2 de este trabajo de grado, se presentan los fundamentos teóricos relacionados con las series temporales y las diferentes distribuciones estadísticas que se analizaron. Adicionalmente se presenta el estado del arte, de diferentes investigaciones realizadas a nivel mundial.

En el capítulo 3 se presenta la metodología que se planteó y empleó para seleccionar las estaciones utilizadas en este trabajo de grado y el procedimiento para realizar los diferentes ajustes a las distribuciones estadísticas, pruebas de bondad de ajuste, puntuación y *ranking*.

El capítulo 4 muestra los resultados de las pruebas de bondad de ajuste realizadas a las diferentes series de datos y finalmente el capítulo 5 presenta las conclusiones y recomendaciones.

Los anexos se presentan en un documento adicional y contienen todos los registros originales de las estaciones, los datos completados y las pruebas estadísticas realizadas.

1.1. Contexto

La proyección de series temporales de variables hidroclimáticas permite pronosticar magnitudes asociadas a una probabilidad de ocurrencia o periodo de retorno, el cual es un parámetro fundamental en el diseño de cualquier estructura hidráulica. Una de las metodologías que existe para proyectar y hacer pronósticos de valores futuros es mediante la aplicación de modelos probabilísticos conocidos como distribuciones estadísticas, las cuales son funciones matemáticas que dependen de parámetros asociados a los registros de las series de datos. Diversos autores han creado familias de distribuciones estadísticas en diferentes campos del conocimiento, pero en hidrología la literatura con base en investigaciones ha popularizado algunas en función del tipo de variable analizada, es por esto que, para series de datos anuales de valores medios se ha popularizado la distribución normal, y para series de datos anuales de valores extremos se ha popularizado la distribución de Gumbel (Europa) y la distribución Log-Pearson 3 (Norteamérica).

Por otra parte, diferentes investigaciones han encontrado que cada región geográfica del planeta es única dentro de la dinámica de los procesos que intervienen el ciclo hidrológico, lo cual hace que las mejores distribuciones estadísticas para una serie de datos de una variable hidroclimática sea diferente según la ubicación geográfica. En el caso de Colombia no existen investigaciones que recomienden o desaconsejen el uso de una distribución estadística específica, limitando esta recomendación a lo establecido en la literatura de otros países. Por lo anterior el presente trabajo de grado busca encontrar si existen distribuciones estadísticas que tengan mejor comportamiento que las que se recomiendan en la literatura.

Para llevar a cabo el trabajo de grado, se tomaron 20 series de datos de precipitación total anual y 20 series de datos de precipitación máxima en 24 horas en una zona específica del departamento de Cundinamarca y se aplicaron 61 distribuciones estadísticas a cada serie de datos. Luego se realizaron pruebas de bondad de ajuste por Kolmogorov-Smirnov y Anderson-Darling y se estableció un sistema de puntajes de 5 a 1 para los mejores ajustes de cada serie de datos. Finalmente se seleccionaron las mejores distribuciones estadísticas y se compararon con aquellas que son referencia en la literatura.

1.2. Objetivos

- Realizar un estado del arte acerca de las distribuciones estadísticas frecuentemente usadas en hidrología, y también de otras que puedan potencialmente ser usadas en hidrología.
- Realizar un análisis comparativo de distribuciones estadísticas no convencionales para estimar una variable hidrológica.
- Proponer distribuciones estadísticas aplicadas en otros campos del conocimiento para usarlas con variables hidrológicas.

2. MARCO TEÓRICO Y ESTADO DEL ARTE

2.1. Marco Teórico

En el estudio de la hidrología existen diferentes tipos de variables, las cuales varían desde climatológicas, como: precipitación, evapotranspiración, brillo solar, humedad relativa, entre otras. Hidráulicas, como: caudales, profundidades de lámina de agua, velocidades, volúmenes, entre otras y variables relacionadas con la calidad del agua. Se miden con procedimientos en algunos casos sencillos y en otros complejos. Estos datos se requieren para planear el uso del recurso hídrico, su control, y para diseñar estructuras o sistemas que lo gestionen (Stephenson D.J, 1981).

Si las series de datos no están autocorrelacionadas, desde el punto de vista estadístico se consideran series aleatorias y el hidrólogo puede tratar la información como un proceso estocástico, siendo posible estimar la probabilidad de que la variable analizada caiga dentro de un rango definido por medio de las distribuciones estadísticas. Este tipo de tratamiento es apropiado para eventos extremos (Inundaciones o sequías) y para información hidrológica promedio, medida en intervalos grandes de tiempo, como por ejemplo la precipitación anual (Chow V. T., 1993).

2.1.1. Caracterización de series de datos máximos

Históricamente las series de datos extremos pueden ser analizadas de dos maneras: Series máximas anuales (AMS) o series de duración parcial (PDS), (Planos E. O., 2015).

Las series máximas anuales (AMS), incluyen únicamente los valores máximos que ocurren cada año durante el periodo de registro.

Las series de duración parcial (PDS), incluyen todos los valores que ocurren en el periodo de registro y que son mayores al valor de umbral. Usando series de duración parcial para el análisis de frecuencias se producen típicamente valores más altos para una frecuencia específica que utilizando series máximas anuales. La diferencia de los valores es mayor en eventos frecuentes de 2,5 y 10 años y disminuye a medida que la frecuencia aumenta. Para periodos de retorno poco frecuentes 25, 50, 100 y 500 años la diferencia en valores es mínima para ambas series.

La relación entre los periodos de retorno de la serie de duración parcial con respecto a la serie anual está dada por la expresión:

$$T_p = \left[\ln \left(\frac{T_a}{T_a - 1} \right) \right]^{-1}$$

Donde:

T_p es el periodo de retorno para la serie de duración parcial

T_a es el periodo de retorno en la serie anual.

De acuerdo con el manual de carreteras (MOP, 2012) para periodos de retorno pequeños, cuando las series son cortas es apropiado utilizar series de duración parcial. También indica que es deseable tener como mínimo 20 años de registro para hacer un análisis de frecuencia con series de máximos anuales y aconseja usar series de duración parcial en caso de que la longitud del registro sea menor.

2.1.2. Datos faltantes

Teniendo en cuenta que en hidrología se utilizan series continuas es importante completar los datos faltantes de las series de datos de la variable hidrológico-climática que se esté analizando, ya que no siempre los datos están completos debido a aspectos como falta de lecturas o ausencia del aparato de medición durante un tiempo (Monsalve G, 1995). A continuación, se presentan los 5 métodos recomendados por la UNESCO – ROSTLAC en 1982 en el marco del balance hídrico para América del sur.

2.1.2.1. Datos completados por regresión lineal

Este método es ampliamente utilizado y consiste en establecer una correlación lineal entre dos estaciones de las cuales una tiene datos incompletos y la otra contiene la información completa y consistente. La ecuación lineal de dos variables es del tipo:

$$Y = a + b * x$$

Donde:

Y es el valor de la precipitación estimada

X es el valor de la precipitación registrada en la estación patrón

a y b son constantes de la regresión

Para aplicar este método es importante validar la calidad del ajuste, y por lo tanto no se debe usar indiscriminadamente, pues los resultados pueden ser no representativos (ROSTLAC, 1982).

Por esta razón, se utiliza el coeficiente de correlación como una forma de evaluar la calidad del ajuste donde el valor del coeficiente R oscila entre -1 y 1, donde un valor de 0 representa correlación nula y valores de -1 o 1 representan una correlación total. En hidrología se considera que la correlación es buena cuando el valor de R es mayor que 0.8 o menor a -0.8 (Pizarro, Gonzáles, Witterssheim, Saavedra, & Soto, 1993).

2.1.2.2. Datos completados por razones de distancias

Este método se utiliza para estimar datos menores a un año, en zonas planas no montañosas, en donde las estaciones deben tener una distribución espacial de acuerdo con la siguiente figura:

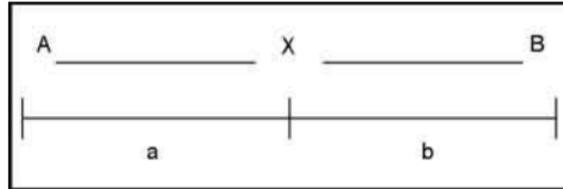


Figura 1 Disposición espacial para completar datos, por razones de distancia. Fuente: Pizarro et al, 1993.

En la figura anterior:

- X representa la posición de la estación que no tiene los datos completos.
- A y B representan las estaciones con información completa.
- a y b son las distancias correspondientes a la estación X.

De acuerdo con el anterior planteamiento se puede calcular el dato faltante mediante la expresión:

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

Donde:

PA, PB y PX representan las precipitaciones para las estaciones A, B y X respectivamente.

Finalmente, este método supone una relación lineal en las precipitaciones, en función de las distancias entre las estaciones.

2.1.2.3. Datos completados por promedios vecinales

En este método se deben tener 3 estaciones con información completa y consistente, en donde la estación con datos incompletos queda al interior de un triángulo proyectado formado por las 3 estaciones patrón.

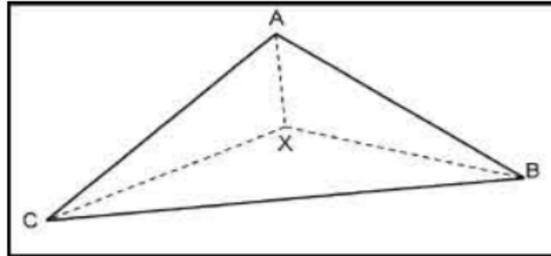


Figura 2 Disposición espacial para completar datos, por promedios vecinales. Fuente: Pizarro et al, 1993.

Donde:

X, A, B y C son las cuatro estaciones pluviométricas y X es la que tiene datos incompletos.

Para estimar el dato faltante, se utiliza la expresión:

$$PX = \sum_{i=1}^n \frac{P_i}{n}$$

Donde:

P_i es la precipitación que corresponde al dato incompleto de la estación X.

n es el número de estaciones

Este método consiste en realizar un promedio aritmético en las n estaciones cercanas existentes que cuentan con información completa, siempre y cuando las precipitaciones de las estaciones vecinas no difieran en más de un 10% unas con respecto a otras.

2.1.2.4. Método de las proporciones

Cuando las precipitaciones de las estaciones patrón difieren en más del 10% del promedio anual multianual es necesario aplicar la siguiente ecuación (ROSTLAC, 1982):

$$P_x = \frac{\bar{P}_x}{N} * \left[\frac{PA}{\bar{P}_A} + \frac{PB}{\bar{P}_B} + \dots + \frac{PN}{\bar{P}_N} \right]$$

Donde; $\bar{P}_x, \bar{P}_A, \bar{P}_B, \bar{P}_N$ = promedio de las precipitaciones anuales registradas en un periodo definido para las N estaciones y la estación X.

PA, PB, PN = Precipitación en las N estaciones durante el periodo que falta X.

Lo que se logra al aplicar este método es estimar el valor faltante, por medio de la relación entre el dato del mes a completar y la precipitación media multianual de cada una de las estaciones patrón (Pizarro et al, 1993).

Para el promedio multianual se deben utilizar los últimos 30 años; de no contarse con este registro se puede establecer un periodo común el cual se sugiere sea mayor en lo posible a 20 años.

2.1.2.5. Método de correlación lineal múltiple

En este método se utilizan las precipitaciones estimadas a partir de correlaciones entre la estación con el dato faltante y cada una de las estaciones vecinas, a las que se asocian los respectivos coeficientes de correlación (ROSTLAC, 1982). La precipitación del dato faltante se estima a partir de la siguiente expresión:

$$P_X = \frac{PXA * rXA + PXB * rXB + \dots + PXN * rXN}{rXA + rXB + \dots + rXN}$$

Donde:

PX es el valor estimado de la precipitación en la estación X la cual tiene el dato faltante

A, B, N son las estaciones consideradas

r es el coeficiente de correlación para cada una de las estaciones.

2.1.3. Distribución empírica de datos

No existe fórmula analítica para determinar la probabilidad de ocurrencia de un evento, pero por lo general se aceptan varias fórmulas empíricas realizadas por diferentes autores. El método de posición de graficación o *plotting position* es una manera de realizar análisis de frecuencias, el cual no se recomienda con registros menores de 10 años.

Con las siguientes ecuaciones se estima lo que se conoce como distribución empírica de la muestra:

Tabla 1 Ecuaciones de probabilidad de excedencia empíricas

Tipo	Año	Ecuación
California	1923	$P = \frac{m}{n}$
Hazen	1930	$P = \frac{2m - 1}{2n}$
Weibull	1939	$P = \frac{m}{n + 1}$
Chegodayev	1955	$P = \frac{m - 0.3}{n + 0.4}$
Blom	1958	$P = \frac{m - 0.375}{n + 0.25}$
Tukey	1962	$P = \frac{3m - 1}{3n + 1}$
Gringorton	1963	$P = \frac{m - 0.44}{n + 0.12}$
Cunnane		$P = \frac{m - 0.4}{n + 0.2}$

Donde:

P es la probabilidad de excedencia

m es el número de orden de los valores o clases de la serie

n es el número de periodos de registro.

2.1.4. Análisis de series de tiempo de variables hidroclimatológicas

Una serie temporal puede ser analizada apropiadamente cuando las características estadísticas básicas son consideradas primero. Varios procedimientos para el análisis de series temporales están basados en hipótesis que deben poseer las series temporales las cuales podrían no ser ciertas. Los resultados de estos análisis basados en falsas hipótesis pueden dar resultados incorrectos o sin representatividad. En el caso particular de la hidrología, este análisis se lleva a cabo principalmente con 4 fines: 1) Construir modelos matemáticos para generar series hidrológicas sintéticas, 2) Predecir eventos hidrológicos, 3) Detectar tendencias y caídas en series hidrológicas y 4) Rellenar datos perdidos y extender series (Alfaro & Soley, 2009).

La mayoría de los análisis de series de tiempo de variables hidroclimatológicas para las escalas de tiempo más comunes, son basados en las siguientes hipótesis: la serie es homogénea, estacionaria, libre de tendencias y no autocorrelacionada (Adeloye & Montaseri, 2002). La homogeneidad y la estacionariedad indican que la serie de datos proviene de una misma población y por lo tanto es estable en media y en varianza. La no homogeneidad y no estacionariedad es debida a diferentes factores como el cambio en el método de la colección de datos, o el cambio en el ambiente en el cual este es realizado, por ejemplo: incendios forestales, derrumbes, explosiones volcánicas, tala indiscriminada de bosques, sobrepastoreo, destrucción de la cobertura vegetal, cambio en el uso del suelo, cambio climático local o global, variabilidad climática, entre otros.

Las pruebas utilizadas para detectar la homogeneidad o consistencia de la serie se dividen en paramétricas y no paramétricas, las pruebas paramétricas asumen que los datos de la serie son normalmente distribuidos (Bethea & Rhinehart, 1994). Las pruebas no paramétricas no asumen ningún origen de la forma en que se distribuyen los datos, por lo cual son más utilizados. Lo anterior no implica que sean mejores, ya que las pruebas paramétricas son más precisas y se ha demostrado que para llegar a la misma conclusión con el mismo nivel de significancia, las pruebas no paramétricas requieren de un 5% a un 35% más datos que las pruebas paramétricas (Bethea & Rhinehart, 1994). Para determinar la estabilidad en la varianza la prueba F o de Fisher es la más utilizada dentro de las pruebas paramétricas y el de Siegel Tukey para las no paramétricas. Detectar cambios en la varianza y en la media son pasos fundamentales para detectar la homogeneidad; se recomienda hacer en primer lugar la prueba para la estabilidad de la varianza, debido a dos razones: la inestabilidad de la varianza implica que la serie de tiempo no es estacionaria y segundo algunas pruebas de estabilidad en la media requieren que haya estabilidad en la varianza (Dahmen & Hall, 1990).

Para establecer la estabilidad en la media de una serie de datos, se recomienda dividirla en dos partes (Maidment, 1997). La prueba t de Student es la más utilizada como prueba paramétrica,

mientras que para las no paramétricas se tienen las pruebas de Mann Withney, Signed Rank, Kruskal-Wallis, Wilcoxon, entre otros.

Por otra parte, la tendencia de una serie de datos puede ser lineal o no lineal, y es importante que no haya tendencias pues la filosofía de la hidrología clásica consiste en considerar que lo que sucedió en el pasado va a suceder en el futuro. Una tendencia puede indicar que hay efectos de cambio climático. La prueba más conocida no paramétrica es la de Mann- Kendall, la cual detecta tendencias monotónicas.

Finalmente, la serie de datos no debe tener autocorrelación, ya que es indeseable debido a que fallan los modelos de distribuciones estadísticas. Además, la autocorrelación indica que los datos no son aleatorios.

2.1.4.1. Prueba F

Esta prueba en honor al estadista británico Ronald Fisher establece la hipótesis nula de que las medias de múltiples poblaciones normalmente distribuidas y con la misma desviación estándar son iguales. Este constituye el problema más simple del análisis de varianza.

En muchos casos, la prueba F se puede resolver explícitamente. Se requieren dos modelos de regresión, de los cuales uno restringe uno o más de los coeficientes de regresión conforme a H_0 . La prueba se basa en un coeficiente modificado de la suma de los cuadrados de residuos, el estadístico es:

$$F = \frac{\left(\frac{RSS_0 - RSS_1}{m}\right)}{\left(\frac{1 - RSS_0}{n - k}\right)}$$

Donde:

RSS_0 se refiere al coeficiente de determinación del modelo sin restringir (R^2)

RSS_1 se refiere al coeficiente de determinación del modelo restringido (R^2)

m se refiere al número de restricciones impuestas a los coeficientes estimados

n es el número de datos observados

El valor crítico del estadístico se obtiene de las tablas, y si F calculado $>$ F tablas; se rechaza la hipótesis nula.

2.1.4.2. Prueba T

Esta prueba paramétrica, asume que los datos son normalmente distribuidos y la hipótesis nula mide si las medias en dos periodos son iguales.

El estadístico de la prueba t se estima a partir de la expresión:

$$t = \frac{(\bar{X} - \bar{Y})}{S \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Donde:

\bar{X} y \bar{Y} son las medias de los primeros y segundos periodos respectivamente

m y n son el número de observaciones en el primer y segundo periodo respectivamente

S es la desviación estándar de la muestra (de las observaciones completas de m y n).

Los valores críticos del estadístico de la prueba para varios niveles de significancia pueden ser obtenidos de las tablas estadísticas de Student's t, y si t calculado > t tablas; se rechaza la hipótesis nula.

2.1.4.3. Prueba de distribución libre CUSUM

Esta prueba no paramétrica fue presentada por McGilchrist y Woodyer en 1975 y asume que la serie de datos tiene una distribución estadística diferente a la normal (distribución libre) y sirve para detectar si las medias de una serie de datos que ha sido dividida en dos partes son significativamente diferentes. La hipótesis nula asume que las medias en dos periodos son iguales, el valor estadístico de la prueba está dado por la expresión:

$$V_k = \sum_{i=1}^k \text{sgn}(X_i - X_{\text{mediana}}) \quad k = 1, 2, 3, \dots, n$$

Donde:

V_k = estadístico de la prueba

$$Sgn(x) = \text{función signo: } \begin{cases} Sgn(x) = 1 \text{ para } X > 0 \\ Sgn(x) = 0 \text{ para } X = 0 \\ Sgn(x) = -1 \text{ para } X < 0 \end{cases}$$

$X_{mediana}$ es la mediana del conjunto de datos de X_i

La distribución de V_k sigue el estadístico de dos muestras de Kolmogorov-Smirnov (KS= (2/n) Max| V_k |) con los valores críticos de Max| V_k | dados por:

$$\alpha = 0.10 \quad 1.22\sqrt{n}$$

$$\alpha = 0.05 \quad 1.36\sqrt{n}$$

$$\alpha = 0.01 \quad 1.63\sqrt{n}$$

Un valor negativo de V_k indica que la última parte de la serie tiene una media mayor que la primera parte y viceversa. Si el valor del estadístico es mayor al estadístico crítico, se rechaza la hipótesis nula.

2.1.4.4. Prueba Rank Sum

La hipótesis nula prueba si las medianas en dos periodos son iguales. Esta es una prueba no paramétrica y para calcular el estadístico de la prueba Rank-Sum se debe seguir el siguiente procedimiento:

Clasificar todos los datos, desde 1 (pequeños) hasta N (grandes). En el caso de las colas (valores de datos iguales), usar el promedio de los rangos.

- Calcular el estadístico S como la suma de los rangos de las observaciones en el grupo pequeño (el número de observaciones en el grupo más pequeño es denotado como n, y el número de observaciones en el grupo grande es denotado como m).
- Calcular la media teórica y la desviación estándar de S bajo la hipótesis nula H_0 para toda la muestra.

$$\mu = n(N + 1)/2$$

$$\sigma = \sqrt{(n * m(N + 1)/12)}$$

La forma estandarizada del estadístico Z_{rs} es calculado como:

$$Z_{rs} = \frac{S - 0.5 - \mu}{\sigma} \text{ si } S > \mu$$

$$Z_{rs} = 0 \text{ si } S = \mu$$

$$Z_{rs} = \frac{|S + 0.5 - \mu|}{\sigma} \text{ si } S < \mu$$

Z_{rs} es aproximadamente distribuido normalmente y el valor crítico del estadístico de la prueba para varios niveles de significancia puede ser obtenido de las tablas de distribución normal.

Si el estadístico de la prueba es mayor al estadístico crítico, se rechaza la hipótesis nula.

2.1.4.5. Prueba de regresión lineal

Esta prueba paramétrica asume que los datos son normalmente distribuidos. Se utiliza para establecer si hay una tendencia lineal, examinando la relación entre el tiempo (x) y la variable de interés (y). La hipótesis nula de la prueba establece que no hay tendencia. El gradiente de regresión se calcula mediante la expresión:

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

La intersección se calcula como:

$$a = \bar{y} * b\bar{x}$$

El estadístico S de la prueba es:

$$S = b/\sigma$$

Donde:

$$\sigma = \sqrt{\frac{12 \sum_{i=1}^n (y_i - a - bx_i)^2}{n(n-2)(n^2-1)}}$$

El estadístico S sigue una distribución t-student con n-2 grados de libertad bajo la hipótesis nula.

Los valores críticos del estadístico para varios niveles de significancia pueden ser obtenidos de la tabla de student. Si el valor del estadístico es mayor al estadístico crítico, se rechaza la hipótesis nula.

La prueba de regresión lineal asume que los datos son normalmente distribuidos y que los errores (desviaciones desde la tendencia) son independientes y siguen la misma distribución normal con media cero.

2.1.4.6. Prueba Mann – Kendall

La prueba de Mann – Kendall se utiliza para establecer si una serie temporal tiene tendencia monótonica ascendente o descendente. Esta prueba no requiere que la serie de datos este normalmente distribuida o sea lineal, pero si requiere que no exista autocorrelación.

La hipótesis nula para esta prueba establece que no existe tendencia, y la hipótesis alterna establece que hay tendencia en la prueba de dos colas o que hay tendencia ascendente o descendente en la prueba de una cola. Para la serie temporal X_1, X_2, \dots, X_n , la prueba MK utiliza el siguiente estadístico:

$$S = \sum_{i=1}^{n-1} \left[\sum_{j=i+1}^n \text{sgn}(R_j - R_i) \right]$$

Donde:

$$\text{Sgn}(x) = \text{función signo: } \begin{cases} \text{Sgn}(x) = 1 \text{ para } X > 0 \\ \text{Sgn}(x) = 0 \text{ para } X = 0 \\ \text{Sgn}(x) = -1 \text{ para } X < 0 \end{cases}$$

Si la hipótesis nula H_0 es verdadera, entonces S se aproximadamente tiene una distribución normal con:

$$\mu = 0$$

$$\sigma = n(n - 1)(2n + 5)/18$$

El estadístico Z es por lo tanto:

$$Z = |S|/\sigma^{0.5}$$

Un valor positivo de S indica que hay una tendencia ascendente y un valor negativo de S indica que hay una tendencia descendente. Los valores críticos del estadístico de la prueba para varios niveles de significancia pueden ser obtenidos de las tablas de distribución normal.

2.1.4.7. Prueba de Autocorrelación

La prueba de autocorrelación por coeficiente de retardo (r1) fue presentada por Jenkins y Watts en 1968. Se utiliza para establecer qué tanta relación existe entre una variable con respecto a si misma en un tiempo pasado de la serie. La hipótesis nula establece que la serie de datos no está autocorrelacionada. El coeficiente de retardo es calculado como:

$$r1 = \frac{\sum_{i=1}^{n-1}(X_i - \bar{X})(X_{i+1} - \bar{X})}{\sum_{i=1}^n(X_i - \bar{X})^2}$$

Si la serie de datos temporal viene de un proceso aleatorio, entonces el valor esperado y la varianza de r1 son:

$$E(r1) = -1/n$$

$$Var(r1) = (n^3 - 3n^2 + 4)/[n^2(n^2 - 1)]$$

El estadístico z es por lo tanto:

$$z = |r1 - E(r1)| / Var(r1)^{0.5}$$

Los valores críticos del estadístico de la prueba para varios niveles de significancia pueden ser obtenidos de las tablas de distribución normal. Si el estadístico de la prueba es mayor al estadístico crítico, se rechaza la hipótesis nula.

2.1.5. Distribuciones de uso común en hidrología

A continuación, se presentan las distribuciones estadísticas más usadas en hidrología, las cuales para la presente tesis de grado corresponden con las presentadas en la literatura de Ven te Chow – Hidrología aplicada:

2.1.5.1. Distribución normal

Es la más común de todas las distribuciones estadísticas, fue presentada inicialmente por el francés Abraham de Moivre en 1733, también es conocida como la distribución Gaussiana, en honor al matemático alemán Carl Friedrich Gauss quien desarrolló en 1809 la función de dos parámetros en estudios de errores en observaciones astronómicas.

La función de densidad está dada por la siguiente expresión:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Donde:

μ = media poblacional

σ = Desviación típica

Si $\mu = 0$ y $\sigma = 1$, se tiene una función de densidad simétrica conocida como normal estándar, la cual se simplifica a la siguiente función:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Al graficar el intervalo de x entre -3.09 y 3.09 se tiene la siguiente figura:

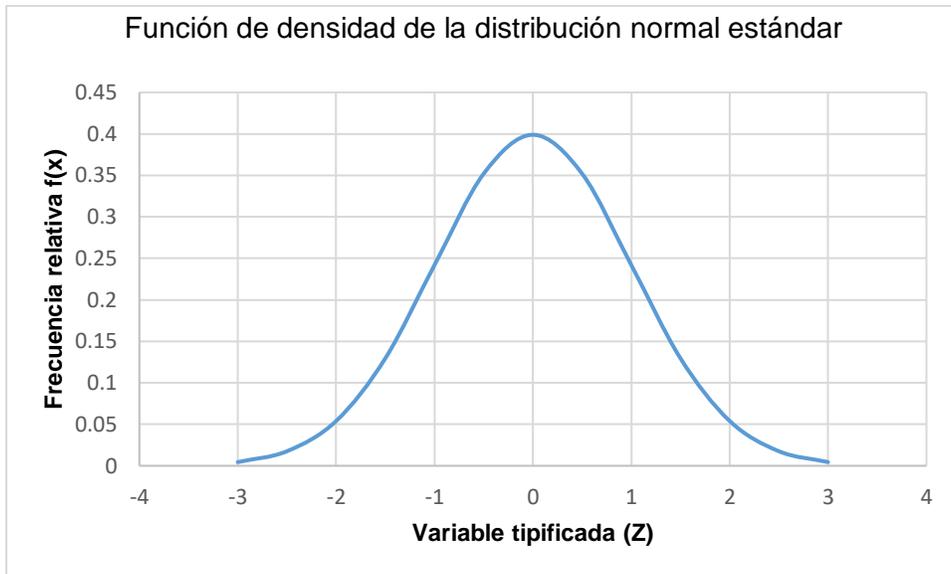


Figura 3 Función de densidad de la distribución normal estándar

La anterior gráfica representa la función $N(0,1)$, con la cual es posible estandarizar o tipificar cualquier función normal a la función normal estándar mediante la expresión.

$$Z = \frac{X - \mu}{\sigma}$$

Donde:

Z es la variable tipificada

X es el valor de $F(x)$

μ es la media de la muestra

σ es la desviación estándar

2.1.5.2. Distribución log normal

Una variable aleatoria X se ajusta a una distribución log normal si la variable $Y = \ln(X)$ se ajusta a una distribución normal. Chow (1954) llegó a la conclusión que esta distribución se aplica a variables hidrológicas que se forman por el producto de varias variables. Por otra parte, se encontró que la distribución log normal funciona bien en la conductividad hidráulica de un medio poroso, los tamaños de las gotas en una tormenta y otras variables hidrológicas (Freeze & Cherry, 1979).

La función de densidad está dada por la siguiente expresión:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$

Donde:

μ = media poblacional

σ = Desviación típica

Esta distribución tiene dos ventajas, la primera es que está limitada a valores de X positivos y la segunda es que disminuye la asimetría positiva común en datos hidrológicos, ya que los logaritmos de los datos más altos se reducen en mayor proporción que los datos más pequeños. Finalmente se requiere que los logaritmos de los datos sean simétricos alrededor de su media.

2.1.5.3. Distribución Exponencial

La distribución exponencial es la equivalente continua de la distribución geométrica discreta. Es útil en procesos en los que se desea conocer el tiempo hasta que ocurra un determinado evento (Procesos de Poisson), teniendo en cuenta que el tiempo que pueda ocurrir en un instante dado t , hasta el tiempo en que este evento ocurra t_i , no depende del tiempo transcurrido anteriormente en el que no pasó nada (Falta de memoria).

Esta distribución es un caso especial de la distribución Gamma ($K=1$) y su función de densidad está dada por la expresión:

$$f(x) = \lambda e^{-\lambda x}$$

Donde:

$$x > 0$$

$$\lambda = 1/\bar{X} \text{ y } \bar{X} \text{ es la media de la muestra}$$

A continuación, se presenta la gráfica de la función de densidad para $\lambda = 1$

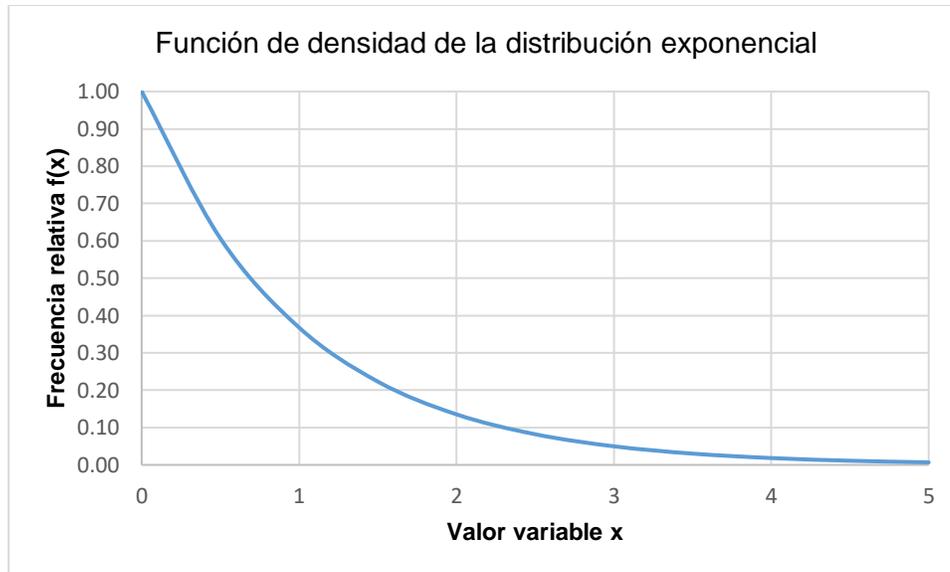


Figura 4 Función de densidad de la distribución exponencial

Algunos procesos hidrológicos se pueden considerar como procesos de Poisson como por ejemplo el volumen de escorrentía contaminada que entra en los ríos a medida que es lavada la superficie de origen de los contaminantes (Chow V. T., 1993). La ventaja de esta distribución, radica en que es fácil obtener el parámetro λ a través de mediciones de campo, y su dificultad está en la hipótesis que cada evento sea independiente de sus vecinos, lo cual podría no ser válido. Muchos investigadores han conseguido procesos de Poisson compuestos donde λ se considera una variable aleatoria y no una constante (Kavvas & Delleur, 1981).

2.1.5.4. Distribución Gamma

La distribución Gamma, es una distribución de probabilidades de dos parámetros λ y β para valores aleatorios mayores o iguales a cero, la cual se utiliza para describir el tiempo hasta que se producen β veces un determinado suceso. La función de densidad está dada por la expresión:

$$f(x) = \frac{\lambda^\beta x^{\beta-1} e^{-\lambda x}}{\Gamma(\beta)}$$

Donde:

Γ = función Gamma

$x > 0$

$\lambda = \frac{\bar{x}}{s_x^2}$ Parámetro de forma

$\beta = \frac{\bar{x}^2}{s_x^2}$ Parámetro de escala

A continuación, se presentan diferentes curvas de densidad de la distribución Gamma, asignando valores a los parámetros de escala y forma (1,1), (1,2) y (1,4) respectivamente:

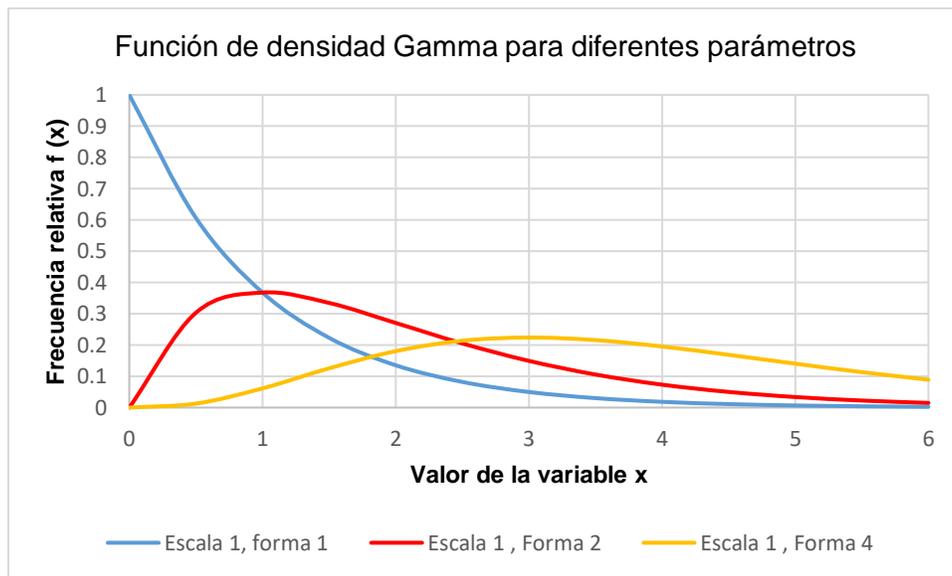


Figura 5 Función de densidad Gamma para diferentes parámetros

La distribución Gamma tiene una forma suave, lo cual la hace conveniente para variables hidrológicas con distribución asimétrica sin el uso de la transformación log (Chow V. T., 1993). Por otra parte, la distribución Gamma de dos parámetros tiene como límite inferior cero, lo cual es una desventaja para las variables hidrológicas que tienen un límite inferior mayor a cero.

2.1.5.5. Distribución Pearson Tipo III

También es conocida como función Gamma de 3 parámetros, fue publicada por primera vez en 1895 por el matemático británico Karl Pearson y se extendió por primera vez a la bioestadística en 1901. La distribución de Pearson tipo III se utilizó por primera vez en la hidrología por Foster en 1924 para describir la distribución de probabilidad de caudales máximos anuales (Chow V. T., 1993).

El sistema de distribuciones de Pearson tiene 7 tipos; todos con solución para $f(x)$ en la siguiente ecuación:

$$\frac{d[f(x)]}{dx} = \frac{f(x)(x - d)}{C_0 + C_1x + C_2x^2}$$

Donde:

d es la moda de la distribución

X es el valor para el cual $f(x)$ es máximo

C_0 , C_1 y C_2 son coeficientes por determinar.

La distribución tipo III de Pearson se logra cuando $C_2 = 0$, y la función de densidad de la distribución de Pearson tipo III está dada por la expresión:

$$f(x) = \frac{\lambda^\beta (x - \epsilon)^{\beta-1} e^{-\lambda(x-\epsilon)}}{\Gamma(\beta)}$$

Donde:

Γ = función Gamma

$$x > \epsilon$$

$$\lambda = \frac{Sx}{\sqrt{\beta}}$$

$$\beta = \left(\frac{z}{Cs}\right)^2$$

$$\epsilon = \bar{X} - Sx\sqrt{\beta}$$

Sx es la desviación estándar de la muestra de datos.

2.1.5.6. Distribución Log Pearson Tipo III

La distribución de log Pearson Tipo III es útil cuando la información presenta un coeficiente de asimetría positivo y se desea reducir este (Chow V. T., 1993). Si log x se ajusta a una distribución Pearson Tipo III, se dice entonces que X se ajusta a log Pearson Tipo III. Esta distribución estadística es la estándar para análisis de caudales máximos en Estados Unidos (Benson, 1968).

La función de densidad está dada por la expresión:

$$f(x) = \frac{\lambda^\beta (y - \epsilon)^{\beta-1} e^{-\lambda(y-\epsilon)}}{x \Gamma(\beta)}$$

Donde:

$$y = \log x$$

Γ = función Gamma

$$\log x > \epsilon$$

$$\lambda = \frac{Sx}{\sqrt{\beta}}$$

$$\beta = \left(\frac{2}{C_s(y)} \right)^2$$

$$\epsilon = \bar{Y} - S_y \sqrt{\beta}$$

En caso tal que la transformación log produzca una curva con asimetría negativa proveniente de información con asimetría positiva, la aplicación de la distribución log Pearson III impondría un límite superior artificial.

2.1.5.7. Distribución de valor extremo

Para cada año de registro histórico disponible de una variable hidroclimatológica se puede conformar una serie de datos de valores extremos, la cual es posible analizar estadísticamente. Se ha demostrado que cualquier serie de datos de valores extremos entre más grande sea, tiende a

converger hacia una de las 3 posibles distribuciones de valores extremos, llamadas tipo I, II y III respectivamente (Fisher & Tippett, 1928).

A continuación, en la Tabla 2 se presentan las funciones de densidad para cada una de las tres distribuciones de valor extremo:

Tabla 2 Funciones de densidad para las distribuciones de valor extremo tipo I (Gumbel), tipo II (Fréchet) y tipo III (Weibull)

Tipo	Autor	Año	Función de densidad
EVI	Gumbel E.J.	1941	$f(x) = \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma} - \exp\left(-\frac{x-\mu}{\sigma}\right)\right)$ <p>Donde:</p> <p>$\sigma > 0$, Parámetro de escala μ = Parámetro de ubicación</p>
EVII	Fréchet M.R.	1927	$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right)$ <p>Donde:</p> <p>$\alpha > 0$, Parámetro de forma $\beta > 0$, Parámetro de escala</p>
EVIII	Weibull E.H.	1939	$f(x) = \frac{\alpha}{\beta} \left(\frac{\beta}{x}\right)^{\alpha+1} \exp\left(-\left(\frac{\beta}{x}\right)^\alpha\right)$ <p>Donde:</p> <p>$\alpha > 0$, Parámetro de forma $\beta > 0$, Parámetro de escala</p>

En 1955 Jenkinson presentó la distribución generalizada de valores extremos (GEV, por sus siglas en inglés), la cual es una función única en la cual las tres funciones limitantes de valores extremos tipo I, II y III son casos especiales.

Esta función fue recomendada por el Natural Environment Research Council en 1975 y es la de mayor aceptación en el Reino Unido para análisis de frecuencias de inundaciones (Ahmad, Sinclair, & Spurr, 1988).

La función de probabilidad de la distribución generalizada de valores extremos (GEV) es:

$$F(x) = \exp\left(-\left(1 - k\frac{x - \mu}{\alpha}\right)^{1/k}\right)$$

Donde μ , k y α son parámetros que deben ser determinados.

Si la variable $k = 0$ la función se limita a la distribución de valor extremo tipo I la cual no tiene límites, para $k < 0$, la función se limita a la distribución de valor extremo tipo II y tiene un límite inferior $(\mu + \frac{\alpha}{k})$ y finalmente si $k > 0$, se limita a la distribución de valor extremo tipo III y tiene un límite superior $(\mu + \frac{\alpha}{k})$.

2.1.6. Distribuciones estadísticas inusuales aplicadas en hidrología

Adicional a las distribuciones estadísticas clásicas en la hidrología presentadas en el numeral anterior, existen diferentes distribuciones estadísticas cuyo uso es inusual en hidrología. Las distribuciones que se presentan a continuación han sido aplicadas en diferentes campos del conocimiento tales como medicina, biología, sociología, econometría, ciencias sociales, acústica, telecomunicaciones, resistencia de materiales, entre otros.

Las distribuciones estadísticas que se estudiaron fueron:

Beta: fue inicialmente discutida por el matemático y bioestadista inglés Karl Pearson en 1895, a la cual el británico William Palin Elderton en 1906 se refería como Pearson Tipo I, su uso se relacionaba con bioestadística y temas financieros.

Burr y Burr 4P: Esta distribución fue introducida en la literatura por Irving Wingate Burr en 1942 (Kumar, 2017) y es conocida como distribución Burr Tipo XII, en el campo de la econometría se conoce como la distribución de Singh – Maddala y se utiliza para modelar ingresos de los hogares.

Cauchy: Es conocida también como la distribución de Cauchy-Lorenz llamada así en honor al matemático francés Augustin Louis Cauchy en 1853 y más adelante por el físico holandés Hendrik

Lorenz. Su uso en la econometría es para calcular rentabilidades económicas (Lehoczky & Schervish, 2018), mientras que en el campo de la física se utiliza para estudiar la oscilación forzada (Resonancia), finalmente en la espectroscopia se aplica para describir la forma de las líneas espectrales.

Chi-Cuadrado y Chi-Cuadrado 2P: Fue presentada por el geodesta Alemán Friedrich Robert Helmert en 1875 (Hald, 1998). Fue redescubierta de manera independiente por el matemático inglés Karl Pearson en el contexto de la bondad de ajuste, con la cual él publicó su prueba de Chi Cuadrado en 1900.

Dagum y Dagum 4P: Fue presentada por el estadista argentino Camilo Dagum en 1970, la distribución de 3 parámetros se conoce como tipo I y la de 4 parámetros como tipo II y se presentó como un modelo en la econometría para estudiar la distribución de la renta.

Erlang y Erlang 3P: Fue presentada por el estadista e ingeniero danés Agner Krarup Erlang en 1909 mientras trabajaba en la compañía de teléfonos de Copenhage, y fue desarrollada para examinar el número de llamadas telefónicas que podían ser realizadas al mismo tiempo por los operadores de estaciones eléctricas, después se extendió a la teoría de colas en la ingeniería de tráfico.

Error: La distribución error, algunas veces llamada distribución generalizada error o distribución generalizada normal son una familia de distribuciones simétricas utilizadas en modelación matemática, usualmente cuando los errores (la diferencia entre el valor esperado y el valor observado) no son distribuidos normalmente. Esta distribución es de especial interés cuando los errores alrededor de la media o de las colas son de especial interés.

Función error: El nombre fue propuesto por el matemático inglés James Whitbread Lee Glaisher en 1871 en su publicación "Teoría de probabilidad, y la teoría de errores". Es importante su utilización para medir la probabilidad de que existan errores en mediciones por ejemplo la tasa de error binario en un sistema digital de comunicaciones.

Exponencial 2P: Es una función especial de la distribución Gamma discutida por Pearson en 1985, en la literatura estadística apareció por primera vez en los estudios de Kondo en 1931, donde se refirió a la distribución exponencial como la distribución de Pearson tipo X. Se ha demostrado su aplicación en problemas actuariales, de biología y de ingeniería (Weibull, 1939). En hidrología la distribución exponencial se aplica para valores extremos de precipitación en 24 h (Ritzema , 1979).

Fatigue life y Fatigue life 3P: También es conocida como la distribución de Birnbaum – Saunders en honor al matemático polaco Zygmunt Wilhelm Birnbaum y por el ingeniero irlandés S.C. Saunders en 1969, como parte de una publicación en el Journal de probabilidad aplicada la cual presentó una nueva familia de distribuciones de dos parámetros para estimar el número de ciclos necesarios para forzar una fisura por fatiga.

Fréchet 3P: Nombrada en honor al matemático francés Maurice Fréchet quien la publicó en 1927, es conocida como Weibull inversa y se utiliza para la estimación de valores extremos.

Gamma 3P: La distribución Gamma es una familia de distribuciones, las cuales se han aplicado en la estimación de lluvias y en contratos de seguros los cuales son modelados como procesos gamma.

Gamma Generalizada, Gamma Generalizada 4P: Esta función es la generalización de la distribución Gamma. Varias distribuciones utilizadas en modelos paramétricos en el análisis de supervivencia tales como la distribución exponencial, la distribución de Weibull y la distribución Gamma son casos especiales de la Gamma generalizada, esta es utilizada frecuentemente para determinar cuál modelo paramétrico es apropiado para una serie de datos.

Pareto Generalizada: Esta familia de distribuciones fue propuesta por Pickands en el año 1975 y surge de la teoría de valores extremos, se utiliza en el estudio de las formas de las colas de una distribución, allí es donde recaen los mayores riesgos en seguros y los valores extremos de precipitación.

Secante hiperbólica: Esta distribución fue presentada por primera vez por Baten W.D. en 1934 y es similar a la normal, está definida por su media y desviación estándar, pero con una curtosis de 5 lo cual representa un pico mayor al de la normal. También se conoce como distribución de coseno hiperbólico inverso.

Gaussiana inversa, Gaussiana inversa 3P: Fue presentada por primera vez en 1900 por el matemático francés Louis Bachelier, y se aplicó para establecer el tiempo en el cual una acción financiera alcanza un máximo por primera vez. Mas adelante fue aplicada de manera independiente por Abraham Wald en 1945 a los procesos de Wiener (Durrett, 2010).

Johnson SB: Presentada en 1949 por el estadista inglés Norman Lloyd Johnson, esta familia de distribuciones de 4 parámetros es versátil, y la sigla SB significa que es la distribución que no tiene límites. Es utilizada en hidrología, meteorología, ingeniería de alimentos, forestal entre otros.

Kumaraswamy: Esta distribución fue presentada en el Journal de hidrología por el hidrólogo hindú Poondi Kumaraswamy en 1980 como una distribución con doble límite. Esta distribución es similar a la Beta, pero mucho más simple de usar especialmente en estudios de simulación ya que la función de densidad y la función de probabilidad acumulada pueden ser expresados de manera explícita.

Laplace: Fue presentada por el matemático francés Pierre Laplace en 1774 y también es conocida como la distribución doble exponencial. Se utiliza en hidrología para estimar valores extremos de precipitación y caudal en ríos (Oosterbaan, 1994) y además para describir épocas de sequía (Burke, Perry, & Brown, 2010).

Levy y Levy 2P: Fue presentada por el matemático francés Paul Levy en el año de 1925, se aplica en espectrometría para describir el perfil de ciertas líneas espectrales donde se conoce como el perfil de van der Waals, también se utiliza en movimiento Browniano, movimiento de precio de acciones e inversión del campo magnético de la tierra (Ahsanullah & Nevzorov, 2014).

Log Gamma: Una variable tiene una distribución log gamma si su logaritmo natural se ajusta a la distribución gamma. Se presentó en 1990 por J. Keiper y es comúnmente usada para modelar datos de pérdidas o severidades en riesgo operativo y seguros (Mora, 2011).

Log Logistic y Log Logistic 3P: Si el logaritmo de una variable aleatoria se ajusta a una distribución Logistic, entonces la serie de datos esta tiene distribución log logistic. Se utiliza en análisis de supervivencia, en hidrología para los caudales y precipitaciones máximos en 24 horas (Shoukri, 1988) en economía para los modelos de distribución de riqueza o ingresos y en redes para modelar los tiempos de transmisión de datos.

Logistic: La función logística fue propuesta inicialmente como una herramienta para estudios demográficos por el matemático Belga Pierre Verhulst entre 1838 y 1847. Se utiliza en hidrología para estimación de caudales medios y lluvias totales mensuales o anuales. Es utilizada por la asociación americana de ajedrez como base para estimar puntajes.

Nakagami: Fue presentada en 1960 por el ingeniero japonés Minoru Nakagami con el propósito de modelar la atenuación de señales y desvanecimiento en comunicaciones inalámbricas.

Pareto y Pareto 2: Fue formulada por el ingeniero francés Vilfredo Pareto en el siglo XIX, tiene bastantes aplicaciones entre las que destacan la distribución del tráfico de internet, valores de reservas en campos de petróleo, tamaño de partículas de arena, tamaño de meteoritos, y en hidrología para estimar eventos extremos de caudal y precipitación (Oosterbaan, 1994) y (Burke, Perry, & Brown, 2010).

Pearson V y Pearson V 3P: La familia de distribuciones de Pearson fue diseñada por el matemático británico Karl Pearson entre 1890 y 1895. Pearson V también es conocida como Gamma Inversa y se utiliza para modelar tiempos de retardo donde hay certeza de un mínimo retardo, pero el máximo no tiene condición de frontera. Por ejemplo, el retardo en los servicios de emergencia o el tiempo para reparar alguna maquinaria (Law & Kelton, 1997).

Pearson VI y Pearson VI 3P: La familia de distribuciones de Pearson fueron diseñadas por el matemático británico Karl Pearson entre 1890 y 1895: También es reconocida como una distribución Beta prima y en general no se conocen varias aplicaciones para esta distribución. Sus tres parámetros brindan flexibilidad, un pico agudo y una cola alargada hacen un posible candidato a datos muy largos que otras distribuciones no ajustan bien.

Pert: Esta distribución fue desarrollada en 1950 por la armada de estados unidos para el sistema armado Polaris, es también conocida como distribución PERT-Beta y su nombre proviene de las redes PERT, las cuales se utilizaban en el pasado para la planeación de proyectos. Se utiliza en planeación de costos o programación de proyectos y sirve para estimar valores más probables entre un valor mínimo y uno máximo.

Power función: Esta función se ajusta bien a series de datos de falla y aporta información apropiada en análisis de confiabilidad y de riesgo. Por ejemplo, los análisis estadísticos de datos obtenidos de dos tipos de ensayos destructivos demuestran la aplicación de la distribución power función.

Rayleigh y Rayleigh 2P: Fue presentada en 1904 por el físico británico John William Strutt (Lord Rayleigh). Su campo de aplicación son las imágenes de resonancias magnéticas y la relación entre la dieta y la respuesta en humanos, también se utiliza para estudiar la altura de las olas en oceanografía.

Recíproca: Es definida como el sinónimo de la distribución logarítmica (Bose & Morin, 2002) y es de especial importancia en el análisis numérico, y en inferencia bayesiana.

Rice: Fue presentada por el ingeniero Americano Stephen O. Rice en 1957 cuando trabajaba en los laboratorios de Nokia. Se aplica en la teoría de la información, en teoría de comunicaciones y en telecomunicaciones.

Student's: Fue presentada por el matemático británico William Sealy Gosset (Student) en el año 1908. Su principal aplicación son los test de student.

Triangular: Fue creada por el matemático británico Thomas Simpson en 1757. Se emplea básicamente para describir de manera subjetiva una población, de la cual solamente se conocen pocos datos muestrales tales como el mínimo, el máximo y el modal. Esta distribución tiene falta de precisión.

Uniforme: Fue creada por el matemático británico Thomas Simpson en 1757. Se emplea en economía para modelar la demanda y reabastecimientos.

A continuación, en la Tabla 3 se presentan las funciones de densidad de todas las distribuciones estadísticas inusuales en hidrología, que se estudiaron en el presente trabajo de grado.

Tabla 3 Distribuciones estadísticas aplicadas en el presente trabajo de grado

Distribución	Función de densidad (PDF)	Rango	Parámetros
Beta	$f(x) = \frac{1}{B(\alpha_1, \alpha_2)} \frac{(x-a)^{\alpha_1-1}(b-x)^{\alpha_2-1}}{(b-a)^{\alpha_1+\alpha_2-1}}$ <p>Donde: B es la función beta</p>	$a \leq x \leq b$	$\alpha_1 > 0$ forma $\alpha_2 > 0$ forma a y b parámetros de limite
Burr	$f(x) = \frac{\alpha k \left(\frac{x}{\beta}\right)^{\alpha-1}}{\beta \left(1 + \left(\frac{x}{\beta}\right)^\alpha\right)^{k+1}}$	$\gamma \leq x \leq \infty$	$k > 0$ forma $\alpha > 0$ forma $\beta > 0$ escala
Burr 4 Páramelos	$f(x) = \frac{\alpha k \left(\frac{x-\gamma}{\beta}\right)^{\alpha-1}}{\beta \left(1 + \left(\frac{x-\gamma}{\beta}\right)^\alpha\right)^{k+1}}$	$\gamma \leq x \leq \infty$	$k > 0$ forma $\alpha > 0$ forma $\beta > 0$ escala $\gamma > 0$ ubicación
Cauchy	$f(x) = \left(\pi\sigma \left(1 + \left(\frac{x-\mu}{\sigma}\right)^2\right)\right)^{-1}$	$-\infty \leq x \leq \infty$	$\sigma > 0$ escala μ Parámetro de limite
Chi cuadrado	$f(x) = \frac{x^{\frac{v}{2}-1} e^{-x/2}}{2^{v/2} \Gamma(\frac{v}{2})}$	$\gamma \leq x \leq \infty$	v Grados de libertad (Entero positivo)

Tabla 3. Continuación

Distribución	Función de densidad (PDF)	Rango	Parámetros
Dagum	$f(x) = \frac{\alpha k \left(\frac{x}{\beta}\right)^{\alpha k - 1}}{\beta \left(1 + \left(\frac{x}{\beta}\right)^\alpha\right)^{k+1}}$	$\gamma \leq x \leq \infty$	$k > 0$ forma $\alpha > 0$ forma $\beta > 0$ escala
Dagum 4 Parámetros	$f(x) = \frac{\alpha k \left(\frac{x - \gamma}{\beta}\right)^{\alpha k - 1}}{\beta \left(1 + \left(\frac{x - \gamma}{\beta}\right)^\alpha\right)^{k+1}}$	$\gamma \leq x \leq \infty$	$k > 0$ forma $\alpha > 0$ forma $\beta > 0$ escala $\gamma > 0$ ubicación
Erlang	$f(x) = \frac{x^{m-1} e^{-x/\beta}}{\beta^m \Gamma(m)}$	$\gamma \leq x \leq \infty$	$m > 0$ forma (entero) $\beta > 0$ escala
Erlang 3 Parámetros	$f(x) = \frac{(x - \gamma)^{m-1} e^{-\frac{x-\gamma}{\beta}}}{\beta^m \Gamma(m)}$	$\gamma \leq x \leq \infty$	$m > 0$ forma (entero) $\beta > 0$ escala $\gamma > 0$ ubicación
Error	$f(x) = C_1 \sigma^{-1} e^{- C_0 z ^k}$ $C_0 = \left(\frac{\Gamma(3/k)}{\Gamma(1/k)}\right)^{1/2}; C_1 = \frac{k C_0}{2\Gamma(1/k)}; z = \frac{x - \mu}{\sigma}$	$-\infty \leq x \leq \infty$	$k > 0$ forma $\sigma > 0$ escala μ Parámetro de limite
Función error	$f(x) = \frac{h}{\sqrt{\pi}} e^{-(hx)^2}$	$-\infty \leq x \leq \infty$	$h > 0$ escala
Exponencial 2 Parámetros	$f(x) = \lambda e^{-\lambda(x-\gamma)}$	$\gamma \leq x \leq \infty$	$\lambda > 0$ escala $\gamma > 0$ ubicación

Tabla 3. Continuación

Distribución	Función de densidad (PDF)	Rango	Parámetros
Distribución de Birnbaum-Saunders (Fatigue life)	$f(x) = \frac{\sqrt{\frac{x}{\beta}} + \sqrt{\frac{\beta}{x}}}{2\alpha x} \cdot \phi\left(\frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)\right)$ <p>ϕ es la función de densidad de la distribución normal estandarizada</p>	$\gamma \leq x \leq \infty$	$\alpha > 0$ forma $\beta > 0$ escala
Distribución de Birnbaum-Saunders (Fatigue life) 3 Parámetros	$f(x) = \frac{\sqrt{\frac{x-\gamma}{\beta}} + \sqrt{\frac{\beta}{x-\gamma}}}{2\alpha(x-\gamma)} \cdot \phi\left(\frac{1}{\alpha}\left(\sqrt{\frac{x-\gamma}{\beta}} - \sqrt{\frac{\beta}{x-\gamma}}\right)\right)$ <p>ϕ es la función de densidad de la distribución normal estandarizada</p>	$\gamma \leq x \leq \infty$	$\alpha > 0$ forma $\beta > 0$ escala $\gamma > 0$ ubicación
Fréchet 3 Parámetros	$f(x) = \frac{\alpha}{\beta} \left(\frac{\beta}{x-\gamma}\right)^{\alpha+1} \cdot e^{-\left(\frac{\beta}{x-\gamma}\right)^\alpha}$	$\gamma \leq x \leq \infty$	$\alpha > 0$ forma $\beta > 0$ escala $\gamma > 0$ ubicación
Gamma 3 Parámetros	$f(x) = \frac{(x-\gamma)^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-(x-\gamma)/\beta}$	$\gamma \leq x \leq \infty$	$\alpha > 0$ forma $\beta > 0$ escala $\gamma > 0$ ubicación
Gamma generalizada	$f(x) = \frac{kx^{k\alpha-1}}{\beta^{k\alpha} \Gamma(\alpha)} \cdot e^{-(x/\beta)^k}$	$\gamma \leq x \leq \infty$	$k > 0$ forma $\alpha > 0$ forma $\beta > 0$ escala

Tabla 3. Continuación

Distribución	Función de densidad (PDF)	Rango	Parámetros
Gamma generalizada Parámetros	4 $f(x) = \frac{k(x-\gamma)^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)} \cdot e^{-((x-\gamma)/\beta)^k}$	$\gamma \leq x < \infty$	$k > 0$ forma $\alpha > 0$ forma $\beta > 0$ escala $\gamma > 0$ ubicación
Pareto generalizada	$f(x) = \begin{cases} \frac{1}{\sigma} \left(1 + k \frac{(x-\mu)}{\sigma}\right)^{-1-\frac{1}{k}} & k \neq 0 \\ \frac{1}{\sigma} e^{-\frac{(x-\mu)}{\sigma}} & k = 0 \end{cases}$	$\mu \leq x < \infty$ para $k \geq 0$ $\mu \leq x \leq \mu - \frac{\sigma}{k}$ para $k < 0$	$k > 0$ forma $\sigma > 0$ escala μ Parámetro de limite
Secante hiperbólica	$f(x) = \frac{\text{Sech}\left(\frac{\pi(x-\mu)}{2\sigma}\right)}{2\sigma}$	$-\infty \leq x < \infty$	$\sigma > 0$ escala μ Parámetro de limite
Gaussiana inversa	$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}}$	$\gamma \leq x < \infty$	$\lambda > 0$ Parámetro μ Parámetro
Gaussiana inversa Parámetros	3 $f(x) = \sqrt{\frac{\lambda}{2\pi(x-\gamma)^3}} e^{-\frac{\lambda(x-\gamma-\mu)^2}{2\mu^2(x-\gamma)}}$	$\gamma \leq x < \infty$	$\lambda > 0$ Parámetro μ Parámetro $\gamma > 0$ ubicación
Johnson SB	$f(x) = \frac{\delta}{\lambda\sqrt{2\pi z(1-z)}} \exp\left(-\frac{1}{2}\left(\gamma + \delta \ln\left(\frac{z}{1-z}\right)\right)^2\right)$ $z = \frac{x-\varepsilon}{\lambda}$	$\varepsilon \leq x \leq \varepsilon + \lambda$	$\gamma > 0$ forma $\delta > 0$ forma $\lambda > 0$ escala $\varepsilon > 0$ ubicación

Tabla 3. Continuación

Distribución	Función de densidad (PDF)	Rango	Parámetros
Kumaraswamy	$f(x) = \frac{\alpha_1 \alpha_2 z^{\alpha_1 - 1} (1 - z^{\alpha_1})^{\alpha_2 - 1}}{(b - a)}$ $z = \frac{x - a}{b - a}$	$a \leq x \leq b$	$\alpha_1 > 0$ forma $\alpha_2 > 0$ forma a y b parámetros de limite; $a < b$
Laplace	$f(x) = \frac{\lambda}{2} e^{-\lambda x-\mu }$	$-\infty \leq x \leq \infty$	$\lambda > 0$ Parámetro μ Parámetro
Levy	$f(x) = \sqrt{\frac{\sigma}{2\pi}} \cdot \frac{e^{-\frac{0.5\sigma}{x}}}{(x - \gamma)^{3/2}}$	$\gamma \leq x \leq \infty$	$\sigma > 0$ escala
Levy 2 Parámetros	$f(x) = \sqrt{\frac{\sigma}{2\pi}} \cdot \frac{e^{-\frac{0.5\sigma}{x-\gamma}}}{(x - \gamma)^{3/2}}$	$\gamma \leq x \leq \infty$	$\sigma > 0$ escala $\gamma > 0$ ubicación
Log Gamma	$f(x) = \frac{(\ln(x))^{\alpha-1}}{x\beta^\alpha \Gamma(\alpha)} \cdot e^{-\ln(x)/\beta}$	$0 \leq x \leq \infty$	$\alpha > 0$ Parámetro $\beta > 0$ Parámetro
Log Logistic	$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot \left(1 + \left(\frac{x}{\beta}\right)^\alpha\right)^{-2}$	$0 \leq x \leq \infty$	$\alpha > 0$ forma $\beta > 0$ escala
Log Logistic 3 Parámetros	$f(x) = \frac{\alpha}{\beta} \left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} \cdot \left(1 + \left(\frac{x-\gamma}{\beta}\right)^\alpha\right)^{-2}$	$0 \leq x \leq \infty$	$\alpha > 0$ forma $\beta > 0$ escala $\gamma > 0$ ubicación

Tabla 3. Continuación

Distribución	Función de densidad (PDF)	Rango	Parámetros
Logistic	$f(x) = \frac{e^{-z}}{\sigma(1 + e^{-z})^2}$ $z = \frac{x - \mu}{\sigma}$	$-\infty \leq x \leq \infty$	$\sigma > 0$ escala $\mu > 0$ ubicación
Lognormal 3 Parámetros	$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2}}{(x-\gamma)\sigma\sqrt{2\pi}}$	$\gamma \leq x \leq \infty$	$\sigma > 0$ parámetro $\mu > 0$ parámetro $\gamma > 0$ ubicación
Nakagami	$f(x) = \frac{2m^m}{\Gamma(m)\Omega^m} x^{2m-1} \cdot e^{-\frac{m}{\Omega}x^2}$	$0 \leq x \leq \infty$	$m \geq 0.5$ parámetro $\Omega > 0$ parámetro
Pareto	$f(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}$	$\beta \leq x \leq \infty$	$\alpha > 0$ forma $\beta > 0$ escala
Pareto 2	$f(x) = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}$	$0 \leq x \leq \infty$	$\alpha > 0$ forma $\beta > 0$ escala
Pearson V	$f(x) = \frac{e^{-\frac{\beta}{x}}}{\beta\Gamma(\alpha)(x/\beta)^{\alpha+1}}$	$\gamma \leq x \leq \infty$	$\alpha > 0$ forma $\beta > 0$ escala

Tabla 3. Continuación

Distribución	Función de densidad (PDF)	Rango	Parámetros
Pearson V 3 Parámetros	$f(x) = \frac{e^{-\frac{x-\gamma}{\beta}}}{\beta \Gamma(\alpha) ((x-\gamma)/\beta)^{\alpha-1}}$	$\gamma \leq x < \infty$	$\alpha > 0$ forma $\beta > 0$ escala $\gamma > 0$ ubicación
Pearson VI	$f(x) = \frac{(x/\beta)^{\alpha_1-1}}{\beta B(\alpha_1, \alpha_2) (1+x/\beta)^{\alpha_1+\alpha_2}}$	$\gamma \leq x < \infty$	$\alpha_1 > 0$ forma $\alpha_2 > 0$ forma $\beta > 0$ escala
Pearson VI 4 Parámetros	$f(x) = \frac{((x-\gamma)/\beta)^{\alpha_1-1}}{\beta B(\alpha_1, \alpha_2) (1+(x-\gamma)/\beta)^{\alpha_1+\alpha_2}}$	$\gamma \leq x < \infty$	$\alpha_1 > 0$ forma $\alpha_2 > 0$ forma $\beta > 0$ escala $\gamma > 0$ ubicación
Pert	$f(x) = \frac{1}{B(\alpha_1, \alpha_2)} \frac{(x-a)^{\alpha_1-1} (b-x)^{\alpha_2-1}}{(b-a)^{\alpha_1+\alpha_2-1}}$ $\alpha_1 = \frac{4m+b-5a}{b-a}$ $\alpha_2 = \frac{5b-a-4m}{b-a}$	$a \leq x \leq b$	m parámetro de moda $a \leq m \leq b$ a y b parámetros de limite; $a < b$
Power Función	$f(x) = \frac{\alpha(x-a)^{\alpha-1}}{(b-a)^\alpha}$	$a \leq x \leq b$	$\alpha > 0$ forma a y b parámetros de limite; $a < b$
Rayleigh	$f(x) = \frac{x}{\sigma^2} \cdot e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$	$\gamma \leq x < \infty$	$\sigma > 0$ escala

Tabla 3. Continuación

Distribución	Función de densidad (PDF)	Rango	Parámetros
Rayleigh 2 Parámetros	$f(x) = \frac{x - \gamma}{\sigma^2} \cdot e^{-\frac{1}{2}\left(\frac{x-\gamma}{\sigma}\right)^2}$	$\gamma \leq x \leq \infty$	$\sigma > 0$ escala $\gamma > 0$ ubicación
Reciproca	$f(x) = \frac{1}{x(\ln(b) - \ln(a))}$	$a \leq x \leq b$	a y b parámetros de limite; $0 < a < b$
Rice	$f(x) = \frac{x}{\sigma^2} \cdot e^{\left(\frac{-(x^2+v^2)}{2\sigma^2}\right)} \cdot I_0\left(\frac{xv}{\sigma^2}\right)$ I_0 es la función modificada de Bessel de primera especie con orden cero	$0 \leq x \leq \infty$	$v \geq 0$ parámetro $\sigma > 0$ parámetro
Student's t	$f(x) = \frac{1}{\sqrt{\pi v}} \frac{\Gamma((v+1)/2)}{\Gamma(v/2)} \left(\frac{v}{v+x^2}\right)^{\frac{v+1}{2}}$	$-\infty \leq x \leq \infty$	v Grados de libertad (Entero positivo)
Triangular	$f(x) = \begin{cases} \frac{2(x-a)}{(m-a)(b-a)} & a \leq x \leq m \\ \frac{2(b-x)}{(b-m)(b-a)} & m \leq x \leq b \end{cases}$	$a \leq x \leq b$	m parámetro de moda $a \leq m \leq b$ a y b parámetros de limite; $a < b$

Tabla 3. Continuación

Distribución	Función de densidad (PDF)	Rango	Parámetros
Uniforme	$f(x) = \frac{1}{b - a}$	$a \leq x \leq b$	a y b parámetros de limite; $a < b$
Weibull 3 Parámetros	$f(x) = \frac{\alpha (x - \gamma)^{\alpha-1}}{\beta^\alpha} e^{-\left(\frac{x-\gamma}{\beta}\right)^\alpha}$	$\gamma \leq x \leq \infty$	$\alpha > 0$ forma $\beta > 0$ escala $\gamma > 0$ ubicación

2.1.7. Métodos de ajuste a una distribución de probabilidad

Una distribución estadística es una función que establece la probabilidad de que una variable aleatoria x sea igualada o excedida. Toda la información de la serie de datos de la variable se puede expresar mediante la función de probabilidad y el valor de los parámetros de la distribución. Para el presente trabajo de grado, se explican el método de los momentos (MOM, por sus siglas en inglés) y el método de la máxima verosimilitud (MLE, por sus siglas en inglés), teniendo en cuenta que son los utilizados por el software EasyFit 5.6, el cual es empleado para realizar los ajustes.

2.1.7.1. Método de los momentos

Este método fue desarrollado en 1902 por Karl Pearson. Se utiliza para determinar el valor de los parámetros de una distribución estadística, igualando el valor de los momentos de la serie de datos, con los momentos de la función de probabilidad.

Si se multiplica la frecuencia relativa de ocurrencia ($1/n$) por el brazo del momento x_1 y se realiza la sumatoria, se obtiene entonces que el momento de primer orden alrededor del origen es la media de la muestra:

$$\sum_{i=1}^n \frac{x_i}{n} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}$$

Teniendo en cuenta que el primer momento es igual al producto del brazo de momento x_1 con la “masa” que en este caso es un diferencial dx de la función de densidad, entonces la media de la muestra es:

$$\mu = \int_{-\infty}^{\infty} xf(x)dx$$

si se desea calcular el valor de un segundo parámetro de la función requerida, se utiliza el momento central de segundo orden el cual es la varianza, $\sigma^2 = E[(x - \mu)^2]$ y para un tercer parámetro, se utiliza el momento central de tercer orden el cual es el coeficiente de asimetría $\gamma = E[(x - \mu)^3]/\sigma^3$.

E es el operador de la esperanza matemática.

2.1.7.2. Método de la máxima verosimilitud

El método de la máxima verosimilitud (MLE, por sus siglas en inglés) fue analizado y popularizado por el biólogo británico R.A Fisher entre 1912 y 1922. La filosofía del método consiste en que el mejor valor para el parámetro de una distribución estadística es aquel que maximiza la verosimilitud o probabilidad conjunta de la serie de datos.

Suponiendo que una serie de datos de una variable aleatoria tiene x_1, x_2, \dots, x_n observaciones independientes e idénticamente distribuidas, la función de densidad para $X = X_i$ es $f(x_i)$, y la probabilidad de que la variable aleatoria ocurra en el intervalo que incluye a x_i es $f(x_i)dx$.

La función de densidad conjunta de todas las observaciones es:

$$f(x_1, x_2, \dots, x_n | \theta) = f(x_1 | \theta) * f(x_2 | \theta) \dots * f(x_n | \theta)$$

La función de verosimilitud supone que mientras los valores de x son fijos, el parámetro θ puede variar, por lo tanto:

$$L(\theta | x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i | \theta)$$

En la práctica es habitual trabajar con la función logaritmo de la verosimilitud, teniendo en cuenta que muchas distribuciones estadísticas son exponenciales (Chow V. T., 1993):

$$\ln L = \sum_{i=1}^n \ln f(x_i | \theta)$$

2.1.8. Pruebas de bondad de ajuste

Las pruebas de bondad de ajuste describen lo bien que se ajusta una serie de datos de una variable aleatoria a una distribución estadística, a través de la medición de la diferencia de los valores observados y los valores esperados según el modelo estadístico.

Existen pruebas que suponen que la serie de datos observada proviene de una distribución estadística estimada, y reciben como nombre pruebas paramétricas. Por otra parte, muchas veces esta suposición no es correcta y se recurre a las pruebas no paramétricas las cuales no hacen ninguna presunción con respecto a la distribución de probabilidad.

Dada una serie de datos independientes (x_1, x_2, \dots, x_n) con distribución F , se desea probar la hipótesis nula $H_0: F=F_0$, donde la hipótesis alternativa es $H_1: F \neq F_0$.

En el presente trabajo de grado se aplican las pruebas de Kolmogorov-Smirnov y Anderson-Darling, ya que son las más utilizadas.

La prueba de Chi-cuadrado no se tuvo en cuenta en este trabajo de grado, ya que fue propuesta por Pearson en 1900 y puede llevar a un resultado sesgado, favoreciendo distribuciones de la familia Pearson (García, 2013).

2.1.8.1. Prueba de Kolmogorov – Smirnov

La prueba de Kolmogorov-Smirnov es una prueba no paramétrica que se emplea para establecer si una serie de datos proviene de una distribución específica.

Las hipótesis a contrastar son:

Ho: (Hipótesis nula) es decir que los datos analizados se ajustan a la distribución estudiada.

H1: (Hipótesis alternativa), es decir que los datos analizados no se ajustan a la distribución estudiada.

El estadístico de contraste es el mayor valor absoluto de la diferencia entre la frecuencia acumulada observada y la frecuencia acumulada teórica.

$$|Dn| = \text{Máx} |Fn(x_i) - Fo(x_i)|$$

Donde:

x_i es el valor i-ésimo observado

$Fn(x_i)$ es el estimador de probabilidad de observar valores menores o iguales a x_i .

$Fo(x_i)$ es la probabilidad de observar valores menores o iguales a x_i , en caso de que Ho sea cierta.

Si $D < D$ de la tabla, se acepta el ajuste con el nivel de significación establecido.

2.1.8.2. Prueba de Anderson-Darling

La prueba de Anderson-Darling es una prueba no paramétrica que se utiliza para comprobar si una serie de datos aleatorios proviene de una distribución específica. Esta prueba da más peso a las colas que la prueba de Kolmogorov-Smirnov (García, 2013).

Si los datos (x_1, x_2, \dots, x_n) los cuales han sido previamente ordenados, provienen de una distribución con función acumulativa F se tiene que las hipótesis de la prueba son:

Ho: (Hipótesis nula) es decir que los datos analizados se ajustan a la distribución estudiada.

H1: (Hipótesis alternativa), es decir que los datos analizados no se ajustan a la distribución estudiada.

Donde el estadístico de la prueba es:

$$A^2 = -N - S$$

El valor de S es:

$$S = \sum_{k=1}^N \frac{2k-1}{N} [\ln F(x) + \ln (1 - F(x_{N+1-k}))]$$

Donde: N es el número de observaciones

Los valores críticos de la prueba de Anderson-Darling dependen de la distribución en particular que se esté estudiando, estos valores han sido formulados y tabulados (Stephens, 1979).

2.1.8.3. Prueba de chi Cuadrado

Esta prueba no paramétrica fue presentada por primera vez en el año 1900 por Pearson y sirve para probar qué tan bien una muestra de datos observados se ajusta a una distribución estadística teórica.

El estadístico de chi cuadrado está definido como:

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Donde k es el número de intervalos, O_i es el número de ocurrencias observadas en el intervalo i y E_i es el número esperado de ocurrencias en el intervalo i.

La función de distribución X^2 se encuentra tabulada en muchos textos de hidrología (Chow V. T., 1993). En la prueba X^2 los grados de libertad están dados por la expresión $v = m - p - 1$

Donde: m es el número de intervalos

p es el número de parámetros utilizados en el ajuste de la distribución propuesta

Si el valor del estadístico es mayor que el valor límite, se rechaza la hipótesis nula H_0 , la cual establece que la distribución propuesta ajusta adecuadamente la información.

2.2. Estado del arte

Muchas distribuciones estadísticas han sido investigadas en la hidrología para el análisis de frecuencias de eventos extremos. A menudo las series de datos de variables hidroclimáticas son asimétricas, lo cual permite usar muchas distribuciones asimétricas, las cuales entre las más comunes que se presentan en la literatura son la Gumbel (EV1), la generalizada de valores extremos (GEV), Log Pearson tipo 3 (LP3) y Log Normal de 3 parámetros (LN3) (Pilon & Harvey, 1994). Cada una de estas distribuciones estadísticas puede demostrar su grado de ajuste, comparando sus resultados teóricos con los valores medidos por las estaciones hidroclimáticas, mediante las pruebas de bondad de ajuste.

Diferentes estudios sobre distribuciones estadísticas sobre series de precipitaciones y registros de escorrentía han sido realizados alrededor del mundo desde el punto de vista de la hidrología clásica. Por ejemplo, en el año de 1965, Markovic realizó en Estados Unidos un estudio con series de datos de 2506 estaciones con mínimo 30 años de registro, encontró que para datos de precipitación el mejor ajuste se logró mediante la distribución log normal de 2 parámetros (LN2) y para caudales anuales se logró el mejor ajuste aplicando la distribución Gamma de 2 parámetros. Se concluyó adicionalmente que las pruebas con dos parámetros requieren menor tiempo de cómputo y dan mejor ajuste que las de 3 parámetros.

(Aguilera, 2007) analizó 28 estaciones y encontró que, para la región de Maule en Chile, las distribuciones estadísticas que tuvieron mejor bondad de ajuste fueron EV1 y Goodrich. (Koschinchanov & Dimitrov, 2009) revisaron las metodologías propuestas para analizar frecuencias de precipitaciones en Bulgaria y encontró que Pearson tipo 3 (P3), regresión polinomial (RP) y LN2 se ajustaron mejor a los datos observados.

En la cuenca del río de las perlas al sur de China, (Fischer, Su, Luo, & Scholten, 2012) encontraron que para precipitaciones máximas en 5 días consecutivos el mejor ajuste es logrado por Wakeby y para precipitaciones máximas anuales el mejor ajuste es logrado por GEV. Por otra parte, en las cuencas de la caída del diablo y Yamaska en la provincia de Quebec en Canadá, para valores de precipitaciones se encontró que el mejor ajuste es logrado por la distribución exponencial mixta (EM) según (Brissette & Chen, 2013). En el mismo año, (Monjo, Caselles, & Chust, 2012) propusieron un modelo alternativo que se ajusta mejor a las series de datos en comparación con las distribuciones estadísticas tradicionales excepto en los meses de verano para España.

Al año siguiente en la región de Badajoz – España, (García, 2013) encontró que las distribuciones estadísticas de Burr, Log logistic de 3 parámetros (LL3), Dagum, GEV y Frechet de 3 parámetros son las que mejor se ajustan a las series de datos de estaciones pluviométricas.

Para la región de Dacca en Bangladés, (Alam, Khan, & Rahat, 2015) realizaron un análisis comparativo con 62 distribuciones estadísticas para dos variables y encontraron que el mejor ajuste logrado para el nivel de lámina de agua del río Dacca se obtuvo con Dagum de 4 parámetros y el caudal máximo anual se ajustó mejor con la distribución de Cauchy.

(Amin, Rizwan, & Alazba, 2015) aplicaron las distribuciones estadísticas Normal, LN2, EV1 y P3 para precipitaciones máximas en 24 horas en el norte de Pakistán y concluyeron que no hay ninguna distribución que sea mejor que otra.

En el norte de China, (Li, Li, Zhao, & Wang, 2015) concluyeron que las distribuciones GEV, Burr y Weibull se ajustan mejor a series de datos de precipitación máxima anual y estacional.

En cuanto a los estudios realizados en África, en Sudán (Mahgoub & Ibrahim, 2015) establecieron que las distribuciones Normal y Gamma son las más recomendadas para las series de datos de precipitación anual con base en la información de 14 estaciones. Por otra parte (Izinyon & Iykeoretin, 2017) realizaron con información de 3 estaciones de Nigeria un estudio sobre distribuciones estadísticas a precipitaciones anuales y concluyeron que la distribución que mejor ajuste tiene es la generalizada de valores extremos (GEV). Finalmente, (Gado & Zeidan, 2018) realizaron un estudio con información de precipitaciones máximas anuales con 30 estaciones y concluyeron que las series se ajustan mejor con las distribuciones estadísticas de LN3 y LP3.

(Mamoon & Rahman, 2017) concluyeron que la distribución GEV es la que mejor se ajusta a series de datos de precipitación máxima en 24 horas en Qatar.

(Thanh, 2017) tomó información de 20 estaciones con series de 30 años y concluyó que, para el área montañosa de Vietnam, los valores extremos de precipitación se ajustan mejor a la distribución LN2, Johnson y LL3.

Adicionalmente, (Alam, Emura, Farnham, & Yuan, 2018) establecieron que para Bangladés las distribuciones GEV, P3 y LP3 son las que mejor se ajustan a los valores extremos de precipitación.

(Esberto, 2018), concluyó en su estudio que, para valores de precipitación anual y mensual en la región de Filipinas, las distribuciones estadísticas que mejor se ajustan son las de Weibull de 3 parámetros, Johnson y LP3.

En Brasil, (Vieira, Casarim, Vismara, & Possenti, 2018) realizaron un estudio con valores de precipitación máxima anual en Paraná y establecieron que la distribución estadística que mejor se ajusta es la de Gamma y Weibull 2 parámetros.

Por otra parte, en el último siglo, uno de los descubrimientos más significativos que contribuyó a la hidrología estadística son los L-momentos de (Hosking, 1990). Las ventajas de los L-momentos son que ellos caracterizan un amplio rango de distribuciones en comparación con los momentos convencionales. También estos son menos sensibles a los valores atípicos. Finalmente, estos se aproximan a su distribución asintótica normal más cercana y por último estos son imparciales para todas las combinaciones de tamaños de muestras y poblaciones (Hosking & Wallis, 1997).

A continuación, en la Tabla 4 se resumen los diferentes estudios de distribuciones estadísticas realizados con base en la hidrología clásica y adicionalmente varios estudios que han utilizado diagramas de L-momentos en el Análisis regional de Frecuencias (ARF):

Tabla 4 Estudios previos sobre distribuciones estadísticas basados en análisis clásicos y análisis regional de frecuencias

Ubicación	Número de estaciones	Distribuciones recomendadas	Referencia
Estados Unidos	2506	LN2 y Gamma 2.	(Markovic, 1965)
Maule, Chile	28	EV1 y Goodrich.	(Aguilera, 2007)
Japón*	22	LP3, GEV, P3, LN3.	(Yue & Hashino, 2007)
Tunisia*	42	Normal y GEV.	(Abida & Ellouze, 2008)
Estados Unidos*	237	P3 y Kappa.	(Hanson & Vogel, 2008)
Bulgaria	7	P3, RP y LN2.	(Koschinchanov & Dimitrov, 2009)
Corea del sur*	37	W3, LN3, GEV, W2 y PG.	(Kim, Lee, Hong, & Lee, 2010)
Sur de China	192	GEV y Wakeby	(Fischer, Su, Luo, & Scholten, 2012)
Quebec, Canadá	24	Exponencial Mixta	(Li, Brissette & Chen, 2012)
España	108	Modelo Alternativo	(Monjo, Caselles, & Chust, 2012)
Badajoz, España	53	Burr, LL3, Dagum, GEV y Frechet.	(García, 2013)
Australia*	127	LP3, GEV y PG.	(Zaman, Rahman, Haddad, & Hagare, 2013)
Daca, Bangladés	5	Dagum 4 y Cauchy.	(Alam, Khan, & Rahat, 2015)
Norte de Pakistán	6	No se recomienda alguna específica.	(Amin, Rizwan, & Alazba, 2015)
Noroccidente de China	13	GEV, Burr y W2	(Li, Li, Zhao, & Wang, 2015)
Sudan	14	Normal y Gamma	(Mahgoub & Ibrahim, 2015)
Sureste de Nigeria	3	GEV	(Izinyon & Iyekeoretin, 2017)
Qatar	29	GEV	(Mamoon & Rahman, 2017)
India*	7	Fisher y Beta.	(Sukrutha & Desai, 2017)

Tabla 4. Continuación

Ubicación	Número de estaciones	Distribuciones recomendadas	Referencia
Sur de Vietnam	20	Johnson, LN2 y LL2.	(Thanh, 2017)
Bangladés	35	GEV, P3 y LP3	(Alam, Emura, Farnham, & Yuan, 2018)
Filipinas	4	W3, Johnson y LP3	(Esberto, 2018)
Egipto	30	LN2 y LP3	(Gado & Zeidan, 2018)
Paraná, Brasil	1	Gamma y W2.	(Vieira, Casarim, Vismara, & Possenti, 2018)

*En estos estudios se aplicó la metodología de los L-momentos con análisis regional de frecuencias (ARF)

(Yue & Hashino, 2007) realizaron un estudio en Japón aplicando L-momentos a series de datos con precipitaciones anuales, estacionales y mensuales a 22 estaciones las cuales tenían 110 años de información cada una y concluyeron que para precipitaciones máximas en 24 horas se lograba un mejor ajuste con LP3 y GEV. Los mejores ajustes para las estaciones eran P3 para la primavera, LP3 para el verano e invierno y LN3 para el otoño. En cuanto a los valores mensuales LP3, LN3 y P3 presentaron los mejores ajustes.

En Tunisia, (Abida & Ellouze, 2008) recomendaron la distribución Normal generalizada y GEV para series de caudales extremos anuales con base en información de 42 estaciones cada una con 27 años de registros. El mismo año (Hanson & Vogel, 2008) concluyeron con base en 237 estaciones que P3 tiene el mejor ajuste en los Estados Unidos de América.

(Kim, Lee, Hong, & Lee, 2010) aplicando ARF en 37 estaciones de corea del sur, concluyeron que los valores extremos de precipitación se ajustan mejor a Weibull 3 parámetros, LN3 y GEV.

(Sukrutha & Desai, 2017) aplicando ARF en 7 estaciones cada una con 100 años de registro en la India, concluyó que las series de precipitación máximas en 24 horas se ajustan mejor a las distribuciones de Fisher y Beta.

Finalmente, en Australia (Zaman, Rahman, Haddad, & Hagare, 2013) concluyeron después de realizar un estudio con diagramas de L-momentos que LP3, GEV y Pareto Generalizada son las que mejor ajuste tienen con series de datos de precipitación máxima en 24 h.

De acuerdo con el análisis de la bibliografía, la distribución estadística que mejor se ajusta a una serie de datos de una variable hidrológica, depende de la región, el tipo y magnitud de la variable y probablemente del ciclo hidrológico.

3. METODOLOGÍA

A continuación, se presenta el procedimiento que se llevó a cabo para realizar el presente trabajo de grado.

3.1. Programas utilizados en el trabajo de grado

En esta sección, se realiza una breve descripción de los programas que se utilizaron para realizar el trabajo de grado:

3.1.1. Trend 1.0.2

Este programa fue desarrollado por *eWater*, que es una organización australiana sin ánimo de lucro. Fue diseñado para detectar tendencias, cambios y aleatoriedad en series de datos. Tiene 12 pruebas estadísticas entre paramétricas y no paramétricas basadas sobre el trabajo “Receta hidrológica: técnicas de estimación de la hidrología australiana” de Grayson et al. para la organización meteorológica mundial y la UNESCO. Este programa se ejecuta en el sistema operativo Windows y es de fácil uso, ya que tiene una interfaz similar a una hoja de cálculo de Excel y adicionalmente es gratuito, principal razón por la cual se seleccionó para utilizarlo en el presente trabajo de grado.

3.1.2. EasyFit 5.6

Este programa fue desarrollado por la empresa ucraniana Mathwave Technologies, la primera versión fue creada en el año 2004. Permite ajustar series de datos a distribuciones estadísticas, seleccionar el mejor modelo, y aplicar los análisis de resultados para tomar decisiones. Se ejecuta en sistema operativo Windows de forma *Stand-alone* o en interfaz compartida con Excel. Es posible con este programa realizar ajustes a más de 50 distribuciones estadísticas discretas de manera automática, seleccionar los mejores ajustes mediante pruebas de bondad de ajuste y generar gráficos interactivos. Este programa se seleccionó teniendo en cuenta que en cuestión de segundos realiza el ajuste de la serie de datos a más de 50 distribuciones estadísticas y adicionalmente tiene una interfaz amigable similar a una hoja de cálculo de Excel. El software EasyFit reporta los valores de los parámetros de las distribuciones estadísticas, los cuales son estimados por los métodos de Máxima verisimilitud (MLE), método de los momentos (MOM) y mínimos cuadrados (LSE).

3.2. Información hidroclimatológica preliminar

Toda la información hidroclimatológica que se consultó en el presente trabajo de grado es la que se encuentra disponible en el sistema de información para la gestión de datos de hidrología y meteorología (DHIME) del IDEAM.

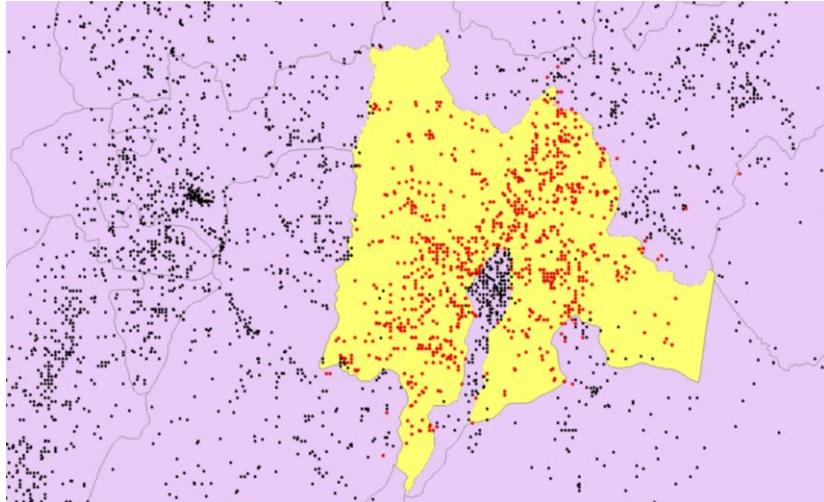


Figura 6 Estaciones hidroclimatológicas en Cundinamarca

En Cundinamarca existen 1051 estaciones hidroclimatológicas, de las cuales 808 se encuentran activas y 243 fueron suspendidas. De estas 808 estaciones activas, 184 fueron instaladas en municipios que tienen como característica el régimen de lluvias de la región Orinoquía (distribución monomodal) y 624 donde el comportamiento del régimen de lluvias es el típico de la región Andina (distribución bimodal).

A continuación, se presenta la Figura 7 donde se observan las categorías de las 624 estaciones activas de Cundinamarca que pertenecen al régimen de lluvias de la región Andina.

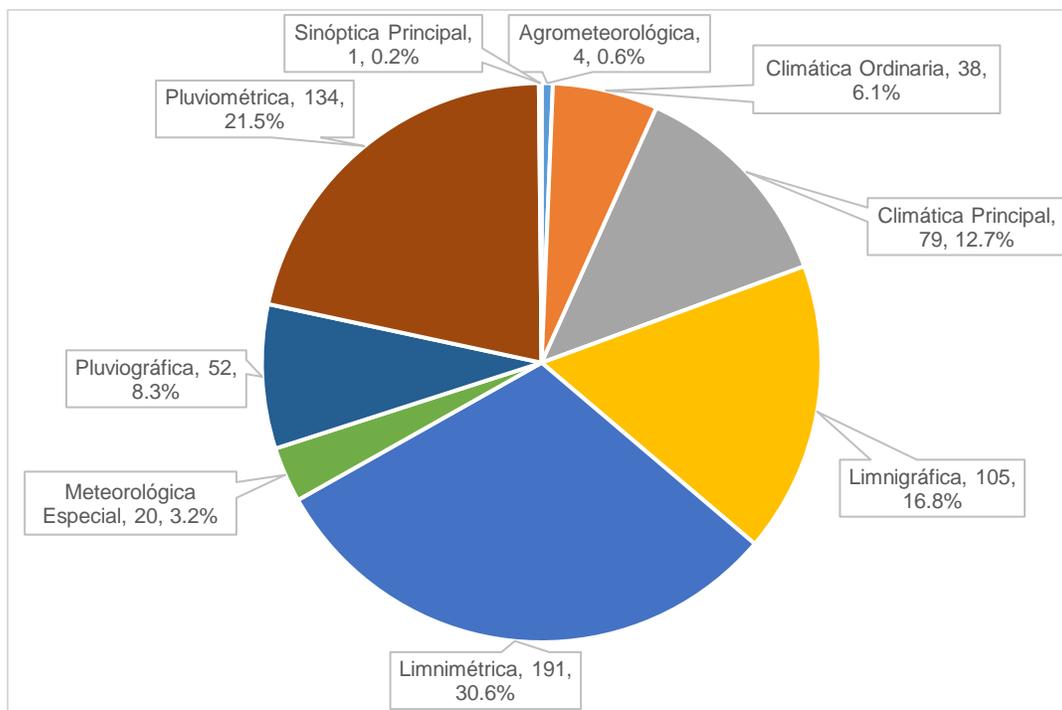


Figura 7 Categoría para las 624 estaciones activas de Cundinamarca en régimen hidrológico Andino

La línea base de las estaciones hidroclimatológicas que se utilizaron en la elaboración del presente trabajo de grado corresponde a las 624 estaciones descritas anteriormente.

3.3. Procedimiento para seleccionar las estaciones hidroclimatológicas

De acuerdo con la línea base de las estaciones que se tuvieron en cuenta para el presente estudio, se definió un procedimiento para seleccionar aquellas estaciones que cumplieron todos los requisitos para poder aplicar sobre sus respectivas series de datos, las distribuciones estadísticas.

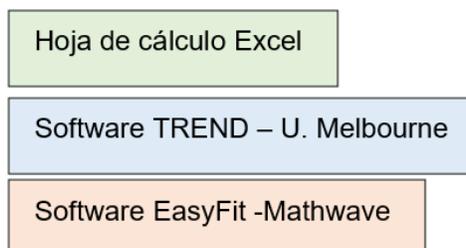
En general el procedimiento que se aplicó fue el siguiente:

1. Se seleccionó una estación hidroclimatológica que tuviera registros mensuales mayores a 30 años de precipitaciones totales mensuales o precipitaciones máximas en 24 horas.
2. Se completaron los datos faltantes utilizando el método de la proporción normal, para lo cual fue necesario identificar una estación testigo cercana a la estación de referencia.
3. Se conformó la serie temporal de valores medios anuales de precipitación o de valores de precipitación máxima anual en 24 horas.

4. Se realizó una prueba de normalidad con significancia del 5% por Kolmogorov-Smirnov a cada serie de datos, para establecer si el análisis posterior se realizaba mediante pruebas paramétricas (distribución normal) o mediante pruebas no paramétricas (distribución no ajustada a la normal).
5. Se realizaron las pruebas de homogeneidad y estacionariedad de la serie de datos, aplicando las pruebas F y t, regresión lineal y autocorrelación para el caso de las pruebas paramétricas y de Cusum, SumRank, Mann-Kendall y autocorrelación en caso de pruebas no paramétricas.
6. En caso de que se aceptaran todas las pruebas estadísticas en la serie de datos, se seleccionó la estación para la siguiente fase y en caso contrario se descartó.
7. Las series de datos de las estaciones seleccionadas se ajustaron a diferentes distribuciones estadísticas mediante el uso del programa EasyFit 5.6, y se estableció un sistema de puntajes de 5 a 1 para cada serie, donde 5 representó la distribución estadística con el mejor ajuste y 1 la distribución estadística con el quinto mejor ajuste. Finalmente se sumaron los puntajes y se realizó el análisis de resultados. Este procedimiento se explica más adelante en la sección 3.6.

A continuación, se presenta el algoritmo que se siguió para establecer la selección o el descarte de una estación hidroclimatológica:

El color que tiene cada proceso del algoritmo representa la herramienta tecnológica que se utilizó para obtener los resultados.



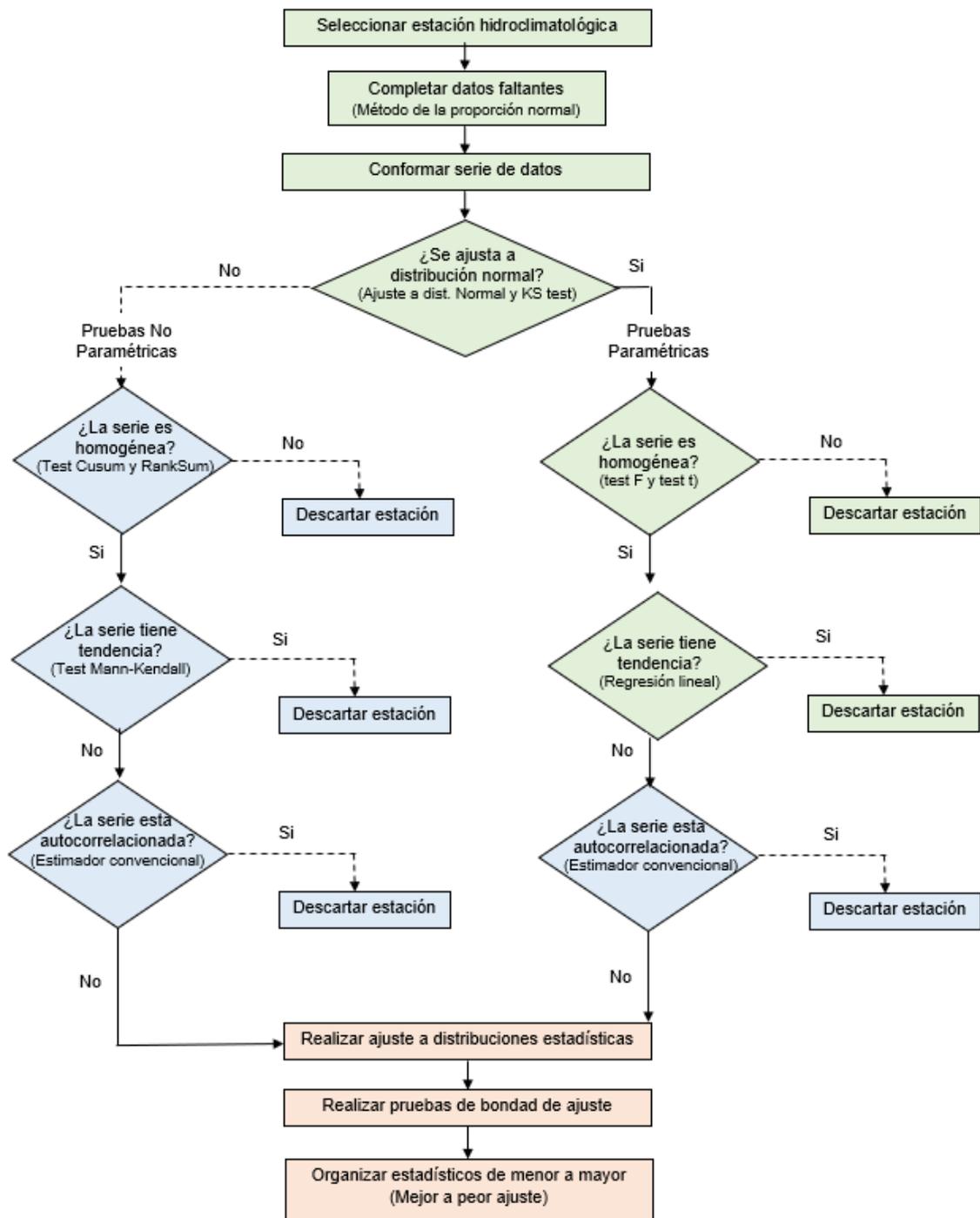


Figura 8 Algoritmo para establecer o descartar una estación hidroclimatológica - fuente:
Elaboración propia

3.3.1. Ejemplo para pruebas paramétricas

A continuación, se presenta un ejemplo de las pruebas estadísticas realizadas a una serie de precipitaciones totales anuales con 37 años de registro, registrados por la estación Núñez (21190330). Se muestran a continuación los datos originales descargados de la página del IDEAM:

Tabla 5 Datos originales precipitaciones totales mensuales, estación Núñez (21190330)

Año/Mes	1	2	3	4	5	6	7	8	9	10	11	12	TOTAL (mm)
1982	14	29	149	241	339	180	71	50	30	85	167	120	1474
1983	107	41	27	109	282	147	41	43	43	54	166	93	1152
1984	108	83	90	71	118	163	158	90	77	139	220	191	1507
1985	97	85	60	131	149	144	36	98	97	134	210	144	1384
1986	12	66	119	51	88	170	84	72	35	74	292	159	1221
1987	57	22	69	206	115	130	81	122	124	107	219	81	1334
1988	34	38	116	103	113	57	165	86	168	117	124	167	1287
1989	110	73	56	138	82	141	87	52	110	100	141	108	1199
1990	60	46	100	91	129	92	35	69	51	69	173	90	1004
1991	145	3	108	413	184	112	97	69	76	48	99	188	1540
1992	92	74	33	61	60	102	49	75	64	70	61	246	987
1993	129	154	75	76	177	182	57	55	82	101	151	115	1353
1994	27	111	80	117	146	139	58	50	75	83	289	183	1358
1995	30	14	96	66	163	125	125	119	111	59	142	110	1160
1996	63	117	78	118	117	118	127	152	100	121	165	117	1393
1997	96	91	38	139	149	115	82	50	31	84	74	167	1115
1998	14	19	103	122	215	213	87	156	73	166	191	182	1540
1999	130	115	181	115	209	106	134	71	151	103	244	136	1695
2000	322	242	195	153	102	208	89	123	68	120	148	148	1918
2001	69	18	141	119	93	172	98	86	55	102	157	194	1305
2002	96	46	60	145	262	122	146	74	60	91	122	82	1305
2003	71	30	80	191	277	134	185	118	87	81	202	105	1562
2004	61	38	152	85	177	174	56	131	44	199	171	212	1500
2005	35	75	73	42	162	170	60	96	62	72	185	113	1145
2006	69	118	45	142	210	200	99	90	90	63	114	163	1405
2007	82	54	48	105	202	193	129	50	121	90	265	116	1456
2008	144	86	95	108	132	196	177	153	118	79	135	198	1622
2009	66	66	111	190	130	63	118	81	69	62	217	135	1308
2010	59	28	45	91	275	206	251	206	123	117	114	235	1751
2011	131	47	147	209	374	166	79	106	81	83	273	186	1881
2012	150	124	63	184	193	139	103	133	106	69	97	86	1446
2013	60	67	84	80	226	160	69	61	76	100	96	0	1078
2014	0	137	122	111	127	170	101	60	42	90	259	152	1369
2015	126	60	64	156	133	52	55	61	40	72	117	122	1060
2016	20	41	55	126	245	0	99	134	62	118	113	126	1138
2017	94	154	126	161	179	209	0	100	51	75	115	159	1423
2018	87	77	203	82	251	204	91	119	76	148	201	133	1674

En color rojo se resaltaron los 4 datos faltantes, los cuales fueron completados empleando como testigo la estación CONC L BUSTAMANTE (21160200) utilizando el método de las proporciones. A continuación, en la Tabla 6 se presentan los datos completos de la estación Núñez:

Tabla 6 Datos completados de precipitaciones totales mensuales, estación Núñez (21190330)

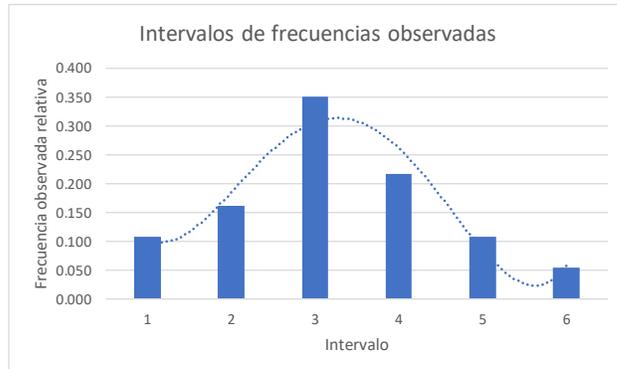
Año/Mes	1	2	3	4	5	6	7	8	9	10	11	12	TOTAL
1982	14	29	149	241	339	180	71	50	30	85	167	120	1474
1983	107	41	27	109	282	147	41	43	43	54	166	93	1152
1984	108	83	90	71	118	163	158	90	77	139	220	191	1507
1985	97	85	60	131	149	144	36	98	97	134	210	144	1384
1986	12	66	119	51	88	170	84	72	35	74	292	159	1221
1987	57	22	69	206	115	130	81	122	124	107	219	81	1334
1988	34	38	116	103	113	57	165	86	168	117	124	167	1287
1989	110	73	56	138	82	141	87	52	110	100	141	108	1199
1990	60	46	100	91	129	92	35	69	51	69	173	90	1004
1991	145	3	108	413	184	112	97	69	76	48	99	188	1540
1992	92	74	33	61	60	102	49	75	64	70	61	246	987
1993	129	154	75	76	177	182	57	55	82	101	151	115	1353
1994	27	111	80	117	146	139	58	50	75	83	289	183	1358
1995	30	14	96	66	163	125	125	119	111	59	142	110	1160
1996	63	117	78	118	117	118	127	152	100	121	165	117	1393
1997	96	91	38	139	149	115	82	50	31	84	74	167	1115
1998	14	19	103	122	215	213	87	156	73	166	191	182	1540
1999	130	115	181	115	209	106	134	71	151	103	244	136	1695
2000	322	242	195	153	102	208	89	123	68	120	148	148	1918
2001	69	18	141	119	93	172	98	86	55	102	157	194	1305
2002	96	46	60	145	262	122	146	74	60	91	122	82	1305
2003	71	30	80	191	277	134	185	118	87	81	202	105	1562
2004	61	38	152	85	177	174	56	131	44	199	171	212	1500
2005	35	75	73	42	162	170	60	96	62	72	185	113	1145
2006	69	118	45	142	210	200	99	90	90	63	114	163	1405
2007	82	54	48	105	202	193	129	50	121	90	265	116	1456
2008	144	86	95	108	132	196	177	153	118	79	135	198	1622
2009	66	66	111	190	130	63	118	81	69	62	217	135	1308
2010	59	28	45	91	275	206	251	206	123	117	114	235	1751
2011	131	47	147	209	374	166	79	106	81	83	273	186	1881
2012	150	124	63	184	193	139	103	133	106	69	97	86	1446
2013	60	67	84	80	226	160	69	61	76	100	96	220	1298
2014	80	137	122	111	127	170	101	60	42	90	259	152	1449
2015	126	60	64	156	133	52	55	61	40	72	117	122	1060
2016	20	41	55	126	245	193	99	134	62	118	113	126	1331
2017	94	154	126	161	179	209	136	100	51	75	115	159	1559
2018	87	77	203	82	251	204	91	119	76	148	201	133	1674

Luego de completar los datos, se conformó la serie de datos anual con las precipitaciones totales y se realizó una prueba de normalidad con una hoja de cálculo en Excel. Se conformaron 6 intervalos y se aplicó la prueba de bondad de ajuste por Kolmogorov – Smirnov con significancia del 5%. Para este ejemplo se tuvo que el valor estadístico de la muestra (0.06) es menor que el valor crítico del estadístico (0.22), por lo tanto, se aceptó la hipótesis nula y se estableció que la serie de datos se ajusta a una distribución normal.

PRUEBA DE BONDAD DE AJUSTE A DISTRIBUCION NORMAL POR KOLMOGOROV SMIRNOV

Media	1396.71
Desviación	223.78
Mínimo	987.37
Máximo	1918.10
Rango	930.73
Número datos	37
Sturges (# intervalos)	6
Raíz N (# intervalos)	6
Tamaño del intervalo	155.12

Intervalos	Límite inferior	Límite superior	Frecuencia Observada	Frecuencia Observada Relativa	Frecuencia Observada Relativa Acumulada	Frecuencia Esperada Relativa Acumulada	ABS (FOR Acum - FER Acum)
1	987.37	1142.49	4	0.108	0.11	0.13	0.02
2	1142.49	1297.61	6	0.162	0.27	0.33	0.06
3	1297.61	1452.73	13	0.351	0.62	0.60	0.02
4	1452.73	1607.86	8	0.216	0.84	0.83	0.01
5	1607.86	1762.98	4	0.108	0.95	0.95	0.00
6	1762.98	1918.10	2	0.054	1.00	0.99	0.01



Estadístico K-S	0.06
Nivel significancia	0.05
Grados de libertad	37.00
Estadístico K-S Crítico	0.22
H0:	Aceptada

Figura 9 Prueba de normalidad en Excel con 6 intervalos mediante Kolmogorov-Smirnov, para la serie de precipitaciones totales anuales, estación Núñez (21190330)

De acuerdo con el algoritmo presentado en la Figura 8, el siguiente paso consiste en realizar la prueba F o de Fisher, para comprobar si al dividir la serie de datos en dos partes iguales, las varianzas de la serie temporal son significativamente diferentes. Esta prueba se realizó utilizando el módulo análisis de datos de Excel. A continuación, en la Tabla 7 se presentan los resultados de la prueba, donde se observa que el P valor es igual a 0.25 y es mayor al nivel de significancia del 5%, para lo cual se concluye que la serie de datos es estacionaria en varianza.

Tabla 7 Prueba F con significancia del 5% para la serie de precipitaciones totales anuales, estación Núñez (21190330)

	Variable 1	Variable 2
Media	1472.31	1316.90
Varianza	52174.34	37666.33
Observaciones	19	18
Grados de libertad	18	17
F	1.38	
P(F<=f) una cola	0.25	
Valor crítico para F (una cola)	2.25	

El siguiente paso consiste en realizar la prueba t o de Student, para comprobar si al dividir la serie de datos en dos partes iguales, las medias de la serie temporal son significativamente diferentes. Esta prueba se realizó utilizando el módulo análisis de datos de Excel. A continuación, en la Tabla 8 se presentan los resultados de la prueba, donde se observa que el P valor es igual a 0.18 y es mayor al nivel de significancia del 5%, para lo cual se concluye que la serie de datos es estacionaria en media.

Tabla 8 Prueba t con significancia del 5% para la serie de precipitaciones totales anuales, estación Núñez (21190330)

	Variable 1	Variable 2
Media	1348.54	1447.54
Varianza	54596.77	42904.32
Observaciones	19	18
Varianza agrupada	48917.58	
Diferencia hipotética de las medias	0	
Grados de libertad	35	
Estadístico t	-1.36	
P(T<=t) una cola	0.09	
Valor crítico de t (una cola)	1.68	
P(T<=t) dos colas	0.18	
Valor crítico de t (dos colas)	2.03	

El siguiente paso corresponde a realizar una regresión lineal, para comprobar si la serie de datos presenta tendencia. Esta prueba se realizó utilizando el módulo análisis de datos de Excel. A continuación, en la Tabla 9 se presentan los resultados de la prueba, donde se observa que el P valor es igual a 0.059 y es mayor al nivel de significancia del 5%, para lo cual se concluye que la serie de datos no tiene tendencia.

Tabla 9 Regresión lineal con significancia del 5% para la serie de precipitaciones totales anuales, estación Núñez (21190330)

Resumen

<i>Estadísticas de la regresión</i>	
Coeficiente de correlación múltiple	0.31
Coeficiente de determinación R ²	0.09
R ² ajustado	0.07
Error típico	215.57
Observaciones	37

ANÁLISIS DE VARIANZA

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	1	176133.91	176133.91	3.78	0.05
Residuos	35	1626580.97	46473.74		
Total	36	1802714.88			

	<i>Coeficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>
Intercepción	-11527.32	6638.75	-1.73	0.09	-25004.71
Variable X 1	6.46	3.31	1.94	0.059	-0.27

Finalmente se realizó la prueba de autocorrelación, aplicando el software TREND V1.0.2, esta prueba permite establecer si los datos provienen o no de un proceso aleatorio.

De acuerdo con los resultados del Software (Ver Figura 10), el valor del estadístico es de 1.20 y el valor crítico para una significancia del 5% es de 1.96, con lo cual se establece que la serie de datos proviene de un proceso aleatorio.

Una vez realizadas estas cuatro pruebas estadísticas, se concluye que la serie de datos es estacionaria en media y en varianza, no tiene tendencia y no está autocorrelacionada, por lo cual se puede proyectar mediante distribuciones estadísticas sin que los resultados presenten errores.

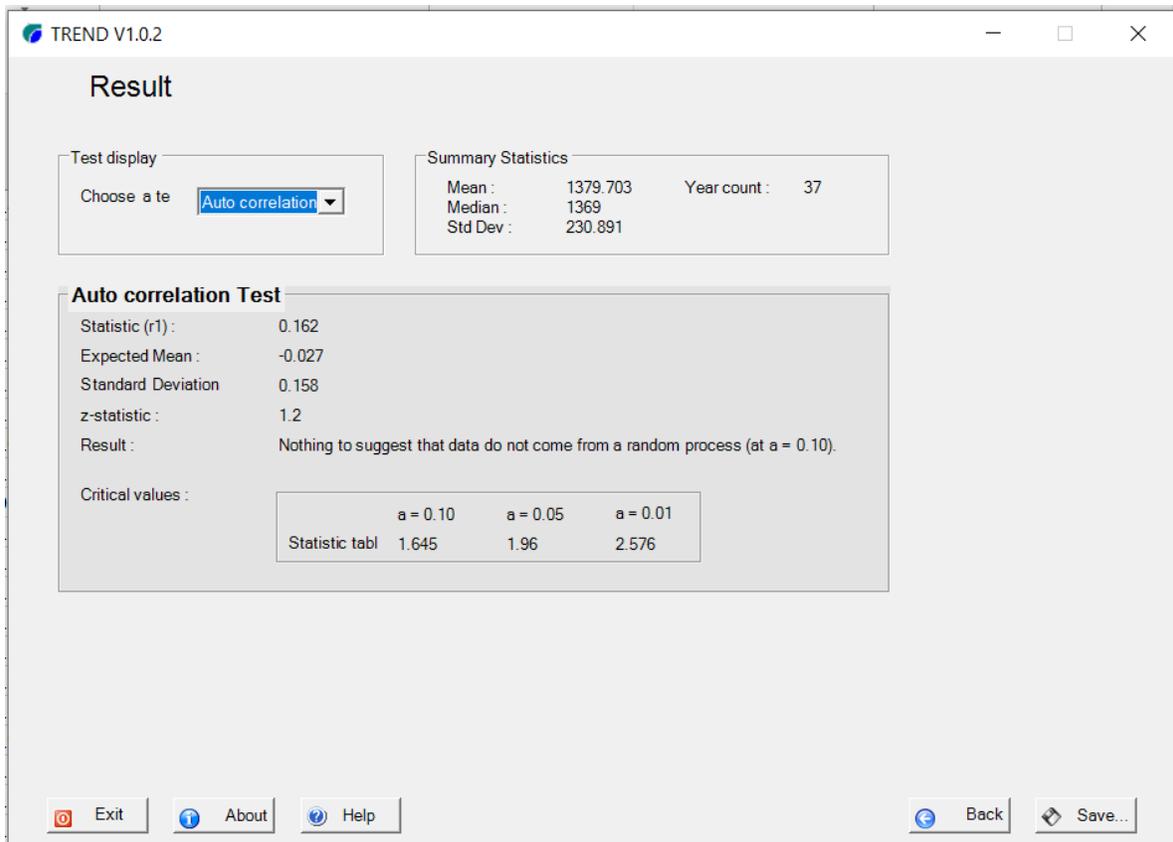


Figura 10 Prueba autocorrelación con software Trend v1.0.2. para la serie de precipitaciones totales anuales, estación Núñez (21190330)

En el documento Anexos, se encuentran todos los registros originales de las estaciones estudiadas, los datos completados y las pruebas estadísticas realizadas.

3.3.2. Ejemplo para pruebas no paramétricas

A continuación, se presenta un ejemplo de las pruebas estadísticas realizadas a una serie anual de precipitaciones máximas en 24 horas con 32 años de registro, que pertenecen a la estación Villa pinzón Granja (21201640). Se muestran a continuación los datos originales descargados de la página del IDEAM:

Tabla 10 Datos originales precipitaciones máximas en 24 horas, estación Villapinzón Granja (21201640)

Año/Mes	1	2	3	4	5	6	7	8	9	10	11	12	MAX (mm)
1988	14.5	9.1	30.1	14.1	19.4	7.6	16.6	6.1	12.1	41.3	27.3	20.5	41.3
1989	13.2	1.7	11.2	34.5	4.5	25.7	13.5	9.2	13.7	12.4	8.1	14.2	34.5
1990	21.3	2.4	17.5	12.2	25.0	26.4	9.2	6.1	9.8	7.2	21.7	15.4	26.4
1991	14.8	7.4	3.8	36.9	14.2	18.5	10.0	19.2	8.1	9.3	23.7	26.1	36.9
1992	25.6	5.2	7.2	11.7	9.4	10.4	8.2	12.9	11.5	15.6	14.0	28.2	28.2
1993	10.5	16.4	6.9	8.8	1.9	12.0	8.8	13.0	3.8	17.7	28.5	31.4	31.4
1994	4.3	7.4	12.9	15.1	11.0	47.9	13.4	9.7	6.0	29.6	2.4	12.0	47.9
1995	2.9	1.7	5.6	25.5	26.5	9.5	31.0	12.2	20.5	9.7	16.3	9.5	31.0
1996	24.9	21.9	15.5	25.6	13.9	47.6	7.5	13.2	10.0	5.7	20.7	19.0	47.6
1997	21.7	17.4	19.2	9.7	10.3	9.8	38.1	8.6	6.4	8.8	19.0	7.2	38.1
1998	1.0	8.7	5.2	13.0	6.8	27.9	10.0	10.0	13.3	15.8	7.2	11.6	27.9
1999	20.5	16.4	24.6	11.7	18.2	16.2	8.5	7.1	4.6	25.0	34.2	25.8	34.2
2000	23.4	9.0	8.6	15.1	13.4	14.9	38.3	16.2	19.5	15.2	41.4	34.7	41.4
2001	10.1	0.9	18.6	4.0	2.5	23.5	7.6	18.9	12.0	9.8	13.4	28.3	28.3
2002	8.7	1.4	12.2	15.3	47.2	32.4	10.8	10.9	14.2	20.6	20.8	9.8	47.2
2003	4.5	0.5	15.0	24.5	14.0	10.9	15.9	11.7	7.4	16.9	17.0	7.5	24.5
2004	3.5	0.2	3.7	40.6	15.6	21.5	14.0	5.4	5.0	5.2	9.0	7.6	40.6
2005	29.4	1.2	13.0	14.0	22.8	15.7	5.9	9.2	9.9	36.6	33.9	24.5	36.6
2006	5.5	14.9	1.0	22.1	32.2	26.2	15.8	10.0	11.4	5.2	15.4	33.1	33.1
2007	17.2	1.2	6.1	8.9	45.3	18.9	10.0	6.9	15.0	6.0	24.5	12.7	45.3
2008	15.8	15.3	38.5	10.3	17.7	21.3	9.9	9.3	29.5	10.5	43.9	25.1	43.9
2009	22.9	27.4	13.6	13.2	14.5	17.8	13.5	10.2	6.6	16.1	17.0	8.3	27.4
2010	2.5	0.0	3.2	10.5	20.2	22.5	21.5	22.2	7.5	15.6	20.8	32.4	32.4
2011	25.8	16.3	22.2	24.5	35.7	54.1	8.1	26.4	13.5	17.9	21.7	29.8	54.1
2012	23.6	24.1	18.1	17.7									24.1
2013		8.5	21.7	22.0	20.3	66.7	4.1	11.4	11.5	14.5	8.9	26.8	66.7
2014	7.9	2.0	22.9	12.5	16.2	10.7	7.5	5.6	6.1	15.0	33.0	14.7	33.0
2015	16.5	12.6	14.2	21.2	4.9	8.1	13.6	8.1	7.7	6.7	11.7	20.0	21.2
2016	6.0	4.2	12.3	17.3	24.0	18.1	5.1	11.5	16.3	9.8	18.0	45.0	45.0
2017	7.6	9.2	19.7	28.0	12.2	31.0	21.9	10.2	6.7	5.8	9.5	16.0	31.0
2018	15.4	12.0	17.3	24.2	29.0	15.9	14.0	11.6	11.3	22.8	9.4	5.2	29.0
2019	0.2	19.3	1.3	15.5	29.2	13.5	8.0	7.0	10.0	18.0	20.3	12.5	29.2

En color rojo y amarillo se resaltaron 10 datos faltantes, los cuales fueron completados empleando como estación testigo VENTAQUEMADA (35070020) utilizando el método de las proporciones. A continuación, en la Tabla 11 se presentan los datos completos de la estación Villapinzón Granja:

Tabla 11 Datos completados precipitaciones máximas en 24 horas, estación Villapinzón Granja (21201640)

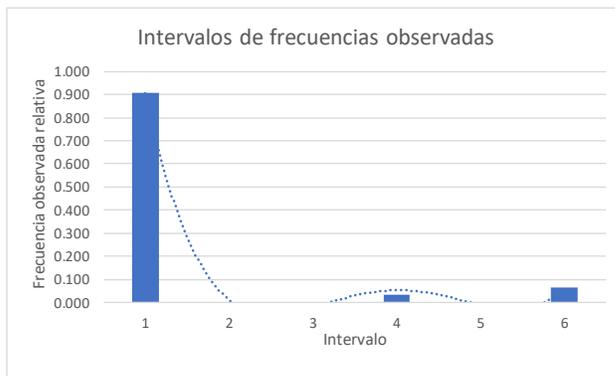
Año/Mes	1	2	3	4	5	6	7	8	9	10	11	12	MAX (mm)
1988	14.5	9.1	30.1	14.1	19.4	7.6	16.6	6.1	12.1	41.3	27.3	20.5	41.3
1989	13.2	1.7	138.3	34.5	4.5	25.7	13.5	9.2	13.7	12.4	310.0	14.2	310.0
1990	21.3	2.4	17.5	12.2	267.6	26.4	9.2	117.0	9.8	7.2	21.7	15.4	267.6
1991	14.8	7.4	3.8	36.9	14.2	18.5	10.0	19.2	8.1	9.3	23.7	26.1	36.9
1992	25.6	5.2	7.2	11.7	9.4	10.4	8.2	12.9	11.5	15.6	14.0	28.2	28.2
1993	10.5	16.4	6.9	8.8	1.9	12.0	8.8	13.0	3.8	17.7	28.5	31.4	31.4
1994	4.3	7.4	12.9	15.1	11.0	47.9	13.4	9.7	6.0	29.6	2.4	12.0	47.9
1995	2.9	1.7	5.6	25.5	26.5	9.5	31.0	12.2	20.5	9.7	16.3	9.5	31.0
1996	24.9	21.9	15.5	25.6	13.9	47.6	7.5	13.2	10.0	5.7	20.7	19.0	47.6
1997	21.7	17.4	19.2	9.7	10.3	9.8	38.1	8.6	6.4	8.8	19.0	7.2	38.1
1998	1.0	8.7	5.2	13.0	6.8	27.9	10.0	10.0	13.3	15.8	7.2	11.6	27.9
1999	20.5	16.4	24.6	11.7	18.2	16.2	8.5	7.1	4.6	25.0	34.2	25.8	34.2
2000	23.4	9.0	8.6	15.1	13.4	14.9	38.3	16.2	19.5	15.2	41.4	34.7	41.4
2001	10.1	0.9	18.6	4.0	2.5	23.5	7.6	18.9	12.0	9.8	13.4	28.3	28.3
2002	8.7	1.4	12.2	15.3	47.2	32.4	10.8	10.9	14.2	20.6	20.8	9.8	47.2
2003	4.5	0.5	15.0	24.5	14.0	10.9	15.9	11.7	7.4	16.9	17.0	7.5	24.5
2004	3.5	0.2	3.7	40.6	15.6	21.5	14.0	5.4	5.0	5.2	9.0	7.6	40.6
2005	29.4	1.2	13.0	14.0	22.8	15.7	5.9	9.2	9.9	36.6	33.9	24.5	36.6
2006	5.5	14.9	1.0	22.1	32.2	26.2	15.8	10.0	11.4	5.2	15.4	33.1	33.1
2007	17.2	1.2	6.1	8.9	45.3	18.9	10.0	6.9	15.0	6.0	24.5	12.7	45.3
2008	15.8	15.3	38.5	10.3	17.7	21.3	9.9	9.3	29.5	10.5	43.9	25.1	43.9
2009	22.9	27.4	13.6	13.2	14.5	17.8	13.5	10.2	6.6	16.1	17.0	8.3	27.4
2010	2.5	1.8	3.2	10.5	20.2	22.5	21.5	22.2	7.5	15.6	20.8	32.4	32.4
2011	25.8	16.3	22.2	24.5	35.7	54.1	8.1	26.4	13.5	17.9	21.7	29.8	54.1
2012	23.6	24.1	18.1	17.7	53.1	9.9	176.3	12.3	7.6	3.6	19.1	4.5	176.3
2013	7.3	8.5	21.7	22.0	20.3	66.7	4.1	11.4	11.5	14.5	8.9	26.8	66.7
2014	7.9	2.0	22.9	12.5	16.2	10.7	7.5	5.6	6.1	15.0	33.0	14.7	33.0
2015	16.5	12.6	14.2	21.2	4.9	8.1	13.6	8.1	7.7	6.7	11.7	20.0	21.2
2016	6.0	4.2	12.3	17.3	24.0	18.1	5.1	11.5	16.3	9.8	18.0	45.0	45.0
2017	7.6	9.2	19.7	28.0	12.2	31.0	21.9	10.2	6.7	5.8	9.5	16.0	31.0
2018	15.4	12.0	17.3	24.2	29.0	15.9	14.0	11.6	11.3	22.8	9.4	5.2	29.0
2019	0.2	19.3	1.3	15.5	29.2	13.5	8.0	7.0	10.0	18.0	20.3	12.5	29.2

Luego de completar los datos, se conformó la serie de datos anual con las precipitaciones máximas en 24 horas y se realizó una prueba de normalidad con una hoja de cálculo en Excel. Se conformaron 6 intervalos y se aplicó la prueba de bondad de ajuste por Kolmogorov – Smirnov con significancia del 5%. Para este ejemplo se tuvo que el valor estadístico de la muestra (0.33) es mayor que el valor crítico (0.24), por lo tanto, se rechazó la hipótesis nula y se estableció que la serie de datos no se ajusta a una distribución normal.

PRUEBA DE BONDAD DE AJUSTE A DISTRIBUCION NORMAL POR KOLMOGOROV SMIRNOV

Media	57.13
Desviación	66.45
Mínimo	21.20
Máximo	310.00
Rango	288.80
Número datos	32
Sturges (# intervalos)	6
Raíz N (# intervalos)	6
Tamaño del intervalo	48.13

Intervalos	Limite inferior	Limite superior	Frecuencia Observada	Frecuencia Observada Relativa	Frecuencia Observada Relativa Acumulada	Frecuencia Esperada Relativa Acumulada	ABS (FOR Acum - FER Acum)
1	21.20	69.33	29	0.906	0.91	0.57	0.33
2	69.33	117.47	0	0.000	0.91	0.82	0.09
3	117.47	165.60	0	0.000	0.91	0.95	0.04
4	165.60	213.73	1	0.031	0.94	0.99	0.05
5	213.73	261.87	0	0.000	0.94	1.00	0.06
6	261.87	310.00	2	0.063	1.00	1.00	0.00



Estadístico K-S	0.33
Nivel significancia	0.05
Grados de libertad	32.00
Estadístico K-S Crítico	0.24
H0:	Rechazada

Figura 11 Prueba de normalidad en Excel, para la serie anual de precipitaciones máximas en 24 horas, estación Villapinzón Granja (21201640)

De acuerdo con el algoritmo presentado en la Figura 8, el siguiente paso consiste en realizar la prueba CuSum, para comprobar si la serie de datos presenta un salto en la varianza. Esta prueba se realizó utilizando el software TREND v1.0.2. A continuación, en la Figura 12 se presenta el resultado de la prueba, donde se observa que la serie de datos es estacionaria en varianza.

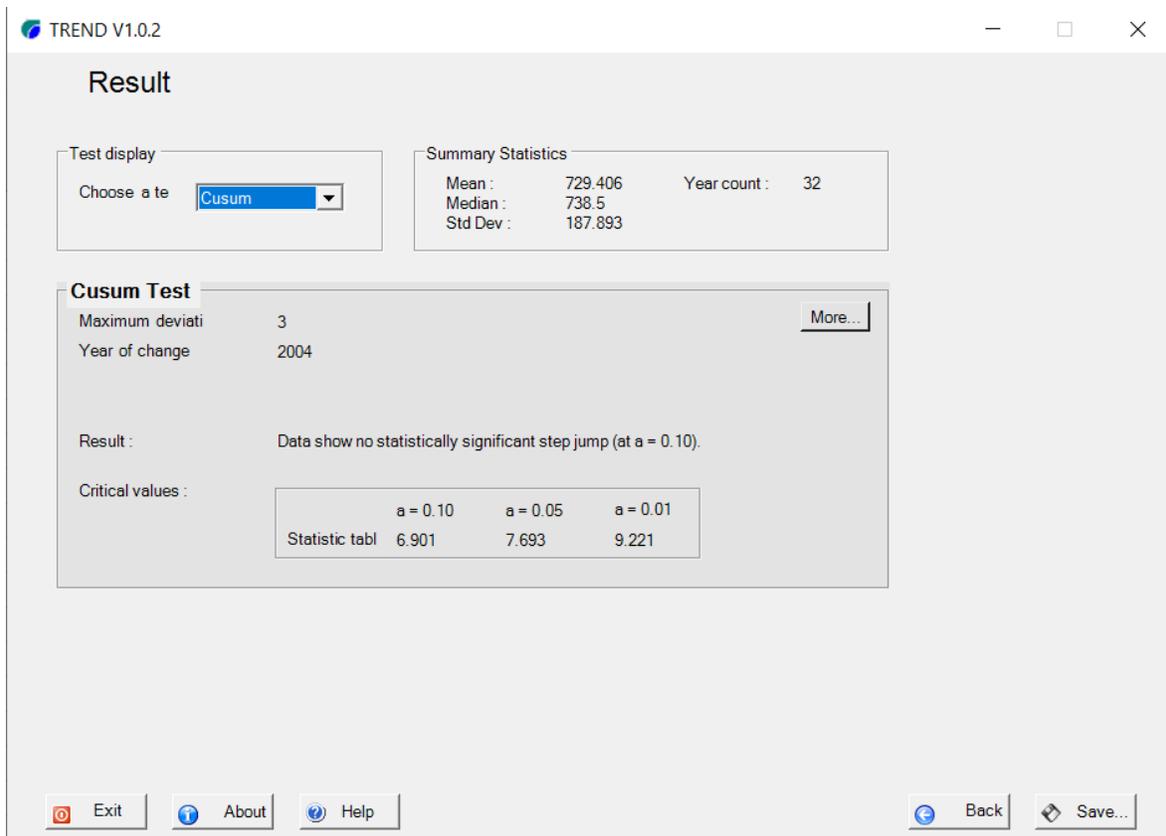


Figura 12 Prueba CuSum con software Trend v1.0.2 para la serie anual de precipitaciones máximas en 24 horas, estación VillaPinzón Granja (21201640)

El siguiente paso consiste en realizar la prueba RankSum, para comprobar si al dividir la serie de datos en dos partes iguales, las medias de la serie temporal son significativamente diferentes. Esta prueba se realizó utilizando el software TREND V1.0.2. A continuación, en la Figura 13 se presentan los resultados de la prueba, donde se observa que el estadístico es igual a -0.66 y es menor al estadístico crítico para un nivel de significancia del 5%, para lo cual se concluye que la serie de datos es estacionaria en media.

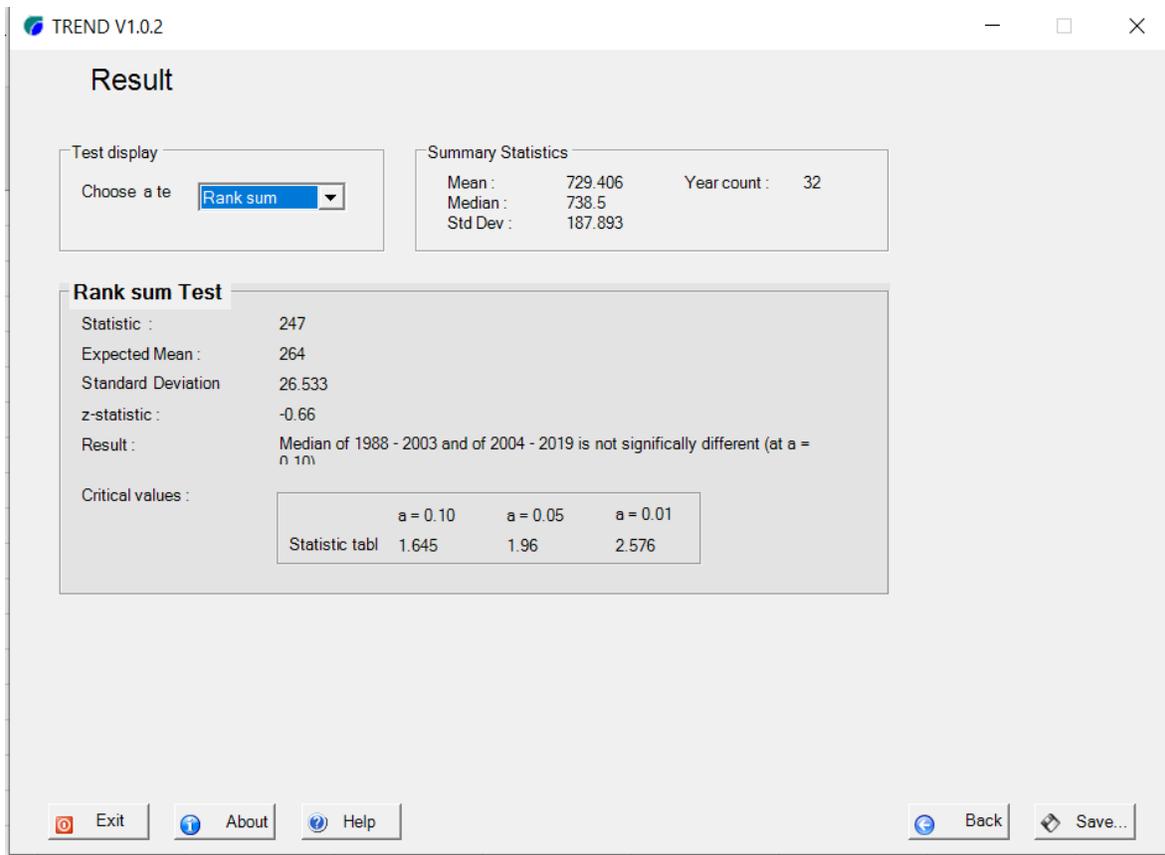


Figura 13 Prueba RankSum con software Trend v.1.0.2 para la serie anual de precipitaciones máximas en 24 horas, estación VillaPinzón Granja (21201640)

El siguiente paso consiste en realizar la prueba de Mann Kendall, para comprobar si la serie de datos presenta tendencia. Esta prueba se realizó utilizando el software TREND v.1.0.2. A continuación, en la Figura 14 se presentan los resultados de la prueba, donde se observa que el valor del estadístico es igual a 0.486 y es menor al valor del estadístico crítico para un nivel de significancia del 5%, para lo cual se concluye que la serie de datos no tiene tendencia.

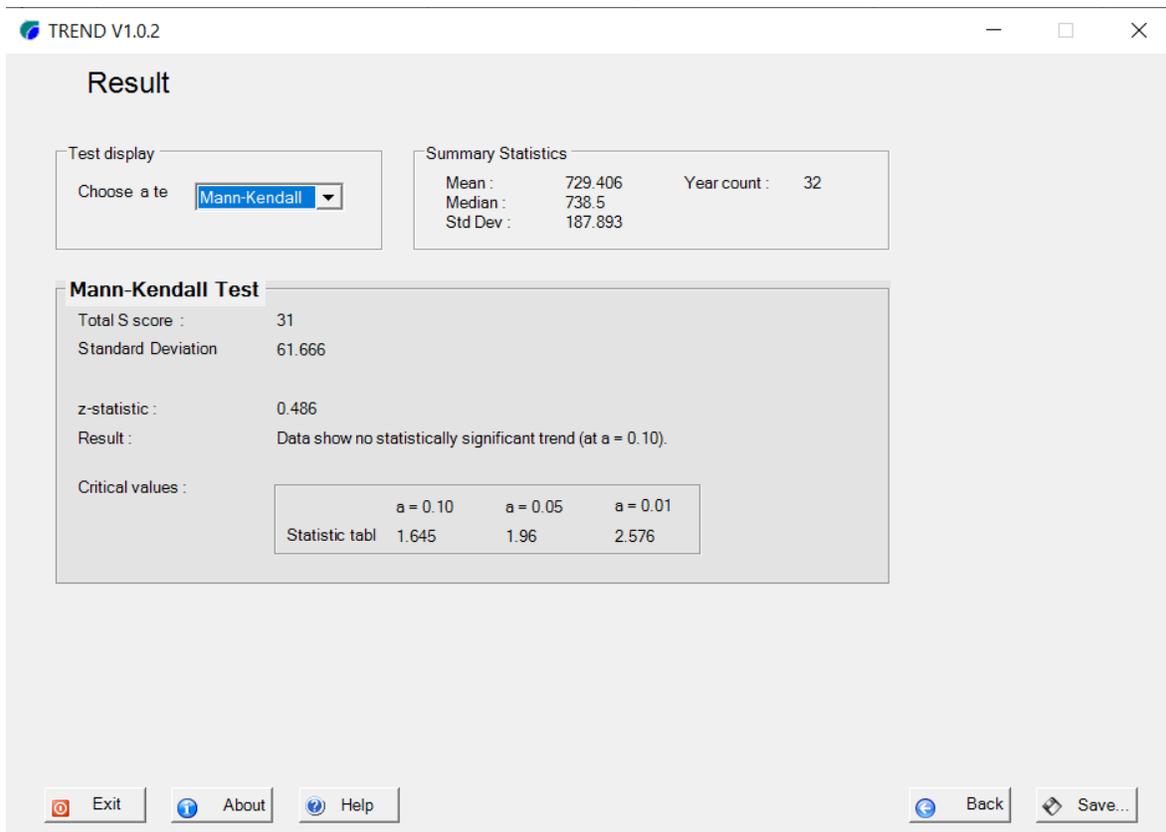


Figura 14 Prueba Mann-Kendall con software Trend v.1.0.2 para la serie anual de precipitaciones máximas en 24 horas, estación VillaPinzón Granja (21201640)

Finalmente se realizó la prueba de autocorrelación, aplicando el software TREND V1.0.2, esta prueba permite establecer si los datos provienen o no de un proceso aleatorio.

De acuerdo con los resultados del Software (Ver Figura 15), el valor del estadístico es de 1.20 y el valor crítico para una significancia del 5% es de 1.96, con lo cual se establece que la serie de datos proviene de un proceso aleatorio.

Una vez realizadas estas cuatro pruebas estadísticas, se concluye que la serie de datos es estacionaria en media y en varianza, no tiene tendencia y no esta autocorrelacionada, por lo cual se puede proyectar mediante distribuciones estadísticas sin que en los resultados se presenten errores.

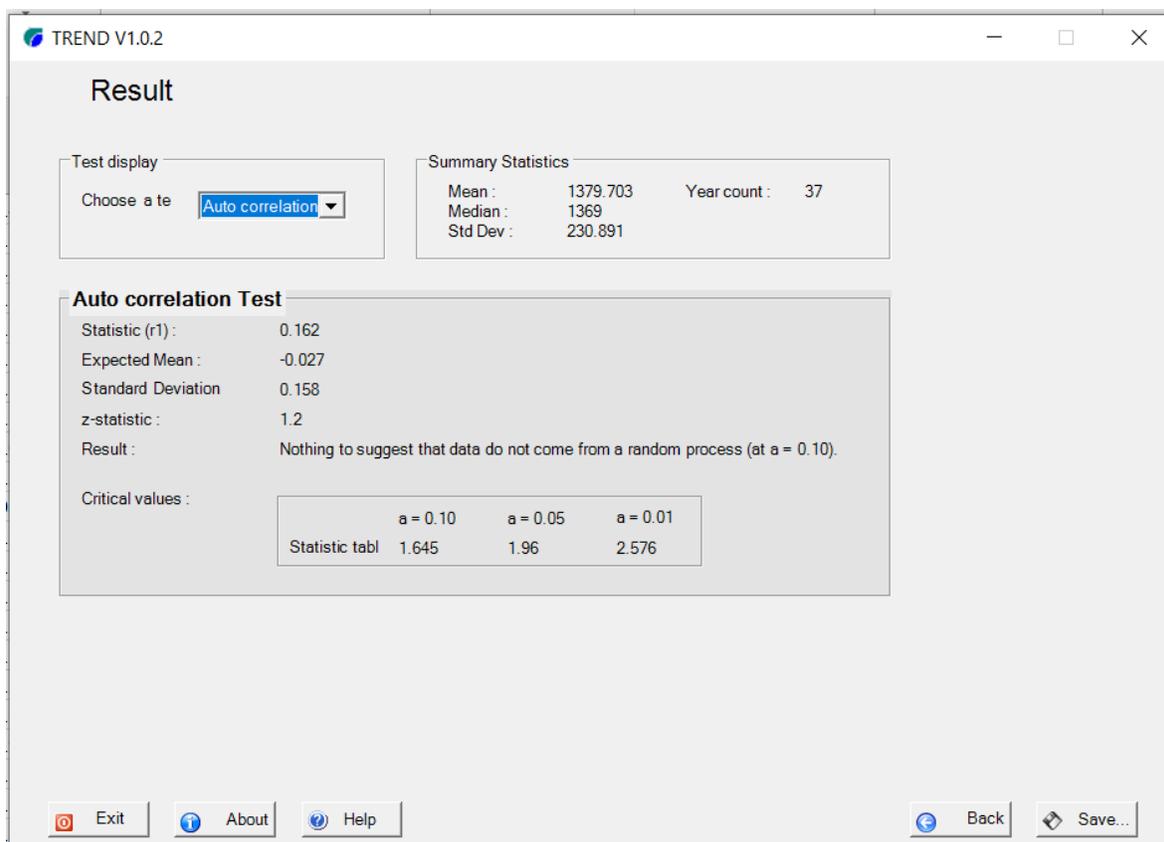


Figura 15 Prueba autocorrelación con software Trend v.1.0.2 para la serie de precipitaciones totales anuales, estación Núñez (21190330)

En el documento Anexos, se encuentran todos los registros originales de las estaciones estudiadas, los datos completados y las pruebas estadísticas realizadas.

3.4. Variables hidroclimatológicas seleccionadas

Teniendo en cuenta la cantidad de estaciones pluviométricas y climatológicas en el departamento de Cundinamarca, y la calidad de la información disponible (homogeneidad, cantidad de años registrados, entre otros), Para desarrollar el presente trabajo de grado, se seleccionaron series anuales de precipitación total y series anuales de precipitación máxima en 24 horas.

3.5. Estaciones empleadas y caracterización del área de estudio

Luego de realizar el procedimiento establecido en la sección anterior a cada una de las estaciones de la línea base, se identificaron 28 estaciones con información de precipitación total anual y 26 estaciones con información de precipitación máxima en 24 horas que cumplieron con las pruebas estadísticas de homogeneidad y estacionariedad.

Finalmente, con el propósito de generar una región de estudio más homogénea y descartar las estaciones que se encontraran en altitudes o presentaran precipitaciones atípicas, se calculó la media y la desviación de las elevaciones y de las precipitaciones y se eliminaron aquellas estaciones que se encontraron por fuera de 1.5 veces la desviación estándar de la muestra. Adicionalmente se eliminaron aquellas estaciones que presentaron una distribución de lluvia anual de tipo monomodal. A continuación, en las Tabla 12 y Tabla 13 se presentan las estaciones descartadas en el último filtro aplicado, en color rojo están resaltados los valores atípicos que corresponden con las estaciones que fueron eliminadas.

Tabla 12 Estaciones descartadas, con altitudes y/o precipitaciones totales por fuera de 1.5 veces la desviación estándar de la muestra para precipitación total anual

Código	Estación	Municipio	Altitud (m.s.n.m)	Años Registro	Precipitación total anual (mm)	Lluvia
21205660	MERCEDES - AUT	Anapoima	810	48	1133.07	Bimodal
21190330	NUNEZ	Cabrera	1950	37	1357.35	Bimodal
24010610	CARMEN DE CARUPA	Carmen De Carupa	2970	35	687.92	Bimodal
23060160	SAN PABLO	Caparrapí	1200	44	2372.91	Bimodal
24010140	CUCUNUBA	Cucunubá	2620	44	723.09	Bimodal
23060180	PENON EL	El Peñón	1400	38	2431.54	Bimodal
21190310	PINAR EL	Fusagasugá	1900	39	1692.79	Bimodal
21205160	RAMADA LA	Funza	2545	30	740.09	Bimodal
23060140	TUSCOLO EL	Guaduas	985	42	1604.06	Bimodal
21205700	GUASCA	Guasca	2750	45	818.00	Bimodal
21200780	POTRERO LARGO	Guatavita	2780	33	1213.20	Monomodal
21201610	SAN ISIDRO	Guasca	2698	32	780.65	Bimodal
21235010	JERUSALEN	Jerusalén	297	44	987.04	Bimodal
23060170	PALMA LA	La Palma	1462	38	2272.40	Bimodal
21201550	ROBLE EL	Madrid	2560	34	856.02	Bimodal
23065120	CABRERA LA	Pacho	1920	45	1530.34	Bimodal
23060260	CHILAGUA FINCA	Nocaima	1500	32	1818.93	Bimodal
21190210	NILO	Nilo	322	35	1535.50	Bimodal
21205720	SAN JORGE GRANJA	Soacha	2900	49	774.51	Bimodal
21201650	SANTA ROSITA	Suesca	2750	30	788.14	Bimodal
23065210	SUPATA	Supatá	1798	45	2005.10	Bimodal
21201210	HATO EL	Tenjo	2575	41	778.28	Bimodal
21201190	LAGUNITAS	Tausa	3100	33	856.32	Bimodal
21205980	PROVIDENCIA GRANJA	Tenjo	2560	36	859.85	Bimodal
21190030	TIBACUY	Tibacuy	1635	33	1164.57	Bimodal
21208900	TOCAIMA	Tocaima	490	44	1128.77	Bimodal
23060190	UTICA	Útica	497	45	1375.32	Bimodal
21201640	VILLAPINZON GRANJA	Villapinzón	2745	32	770.03	Bimodal
		Promedio	1898.83		1301.82	
		Desviación estándar.	885.18		592.87	
		Lim. Superior	3226.59		2191.13	
		Lim. Inferior	571.06		412.51	

Tabla 13 Estaciones descartadas, con altitudes y/o precipitaciones totales por fuera de 1.5 veces la desviación estándar de la muestra para precipitación máxima en 24 horas

Código	Estación	Municipio	Altitud (m.s.n.m)	Años Registro	P máx 24 h (mm)	Lluvia
24015120	ISLA DEL SANTUARIO	Fúquene	2580	32	47.12	Bimodal
21205660	MERCEDES - AUT	Anapoima	810	48	69.71	Bimodal
21190330	NUNEZ	Cabrera	1950	30	48.03	Bimodal
24010610	CARMEN DE CARUPA	Carmen De Carupa	2970	36	36.22	Bimodal
23060160	SAN PABLO	Caparrapí	1200	38	88.63	Bimodal
24010140	CUCUNUBA	Cucunubá	2620	44	43.82	Bimodal
23060180	PENON EL	El Peñón	1400	45	86.20	Bimodal
21190310	PINAR EL	Fusagasugá	1900	30	64.19	Bimodal
23060140	TUSCOLO EL	Guaduas	985	41	85.90	Bimodal
21205700	GUASCA	Guasca	2750	45	38.33	Bimodal
21200780	POTRERO LARGO	Guatavita	2780	33	44.48	Monomodal
21201610	SAN ISIDRO	Guasca	2698	30	33.70	Bimodal
21235010	JERUSALEN	Jerusalén	297	44	72.39	Bimodal
23065120	CABRERA LA	Pacho	1920	30	61.66	Bimodal
23060260	CHILAGUA FINCA	Nocaima	1500	32	74.28	Bimodal
21190210	NILO	Nilo	322	32	97.40	Bimodal
21205720	SAN JORGE GRANJA	Soacha	2900	49	34.75	Bimodal
21205920	SUASUQUE	Sopó	2650	37	40.04	Monomodal
23065210	SUPATA	Supatá	1798	40	75.10	Bimodal
21201210	HATO EL	Tenjo	2575	41	39.46	Bimodal
21201190	LAGUNITAS	Tausa	3100	33	39.97	Bimodal
21190030	TIBACUY	Tibacuy	1635	33	59.69	Bimodal
21208900	TOCAIMA	Tocaima	490	44	77.21	Bimodal
23060190	UTICA	Útica	497	45	77.71	Bimodal
21201640	VILLAPINZON GRANJA	Villapinzón	2745	32	57.13	Bimodal
21201060	PANTANO REDONDO 1	Zipacquirá	3160	30	46.22	Bimodal
		Promedio	1910.33		60.37	
		Desviación estándar	927.10		20.03	
		Límite Superior	3300.98		90.41	
		Límite Inferior	519.69		30.34	

Para la variable de precipitación total anual, se obtuvo un total de 20 estaciones hidroclimatológicas que se emplearon para el análisis comparativo de distribuciones estadísticas del presente trabajo de grado, se definió una caracterización de la zona de estudio, donde las altitudes de las estaciones varían desde los 810 m.s.n.m. hasta los 3100 m.s.n.m. y las precipitaciones totales anuales están en un rango de 688 mm/año hasta los 2431 mm/ año. En todos los casos la distribución de la lluvia anual es de tipo bimodal, lo cual indica que al año se presentan dos picos de meses lluviosos. A continuación, se presentan las estaciones definitivas, que se emplearon en este trabajo de grado:

Tabla 14 Estaciones hidroclimatológicas definitivas para precipitación total anual, empleadas en el presente trabajo de grado

Código	Estación	Municipio	Altitud (m.s.n.m)	Años Registro
21205660	MERCEDES - AUT	Anapoima	810	48
21190330	NUNEZ	Cabrera	1950	37
24010610	CARMEN DE CARUPA	Carmen De Carupa	2970	35
24010140	CUCUNUBA	Cucunubá	2620	44
21190310	PINAR EL	Fusagasugá	1900	39
21205160	RAMADA LA	Funza	2545	30
23060140	TUSCOLO EL	Guaduas	985	42
21205700	GUASCA	Guasca	2750	45
21201610	SAN ISIDRO	Guasca	2698	32
21201550	ROBLE EL	Madrid	2560	34
23065120	CABRERA LA	Pacho	1920	45
23060260	CHILAGUA FINCA	Nocaima	1500	32
21205720	SAN JORGE GRANJA	Soacha	2900	49
21201650	SANTA ROSITA	Suesca	2750	30
23065210	SUPATA	Supatá	1798	45
21201210	HATO EL	Tenjo	2575	41
21201190	LAGUNITAS	Tausa	3100	33
21205980	PROVIDENCIA GRANJA	Tenjo	2560	36
21190030	TIBACUY	Tibacuy	1635	33
21201640	VILLAPINZON GRANJA	Villapinzón	2745	32

Así mismo, se presenta en la Figura 16 el mapa del departamento de Cundinamarca con las estaciones seleccionadas para estudiar la variable de precipitación total anual:

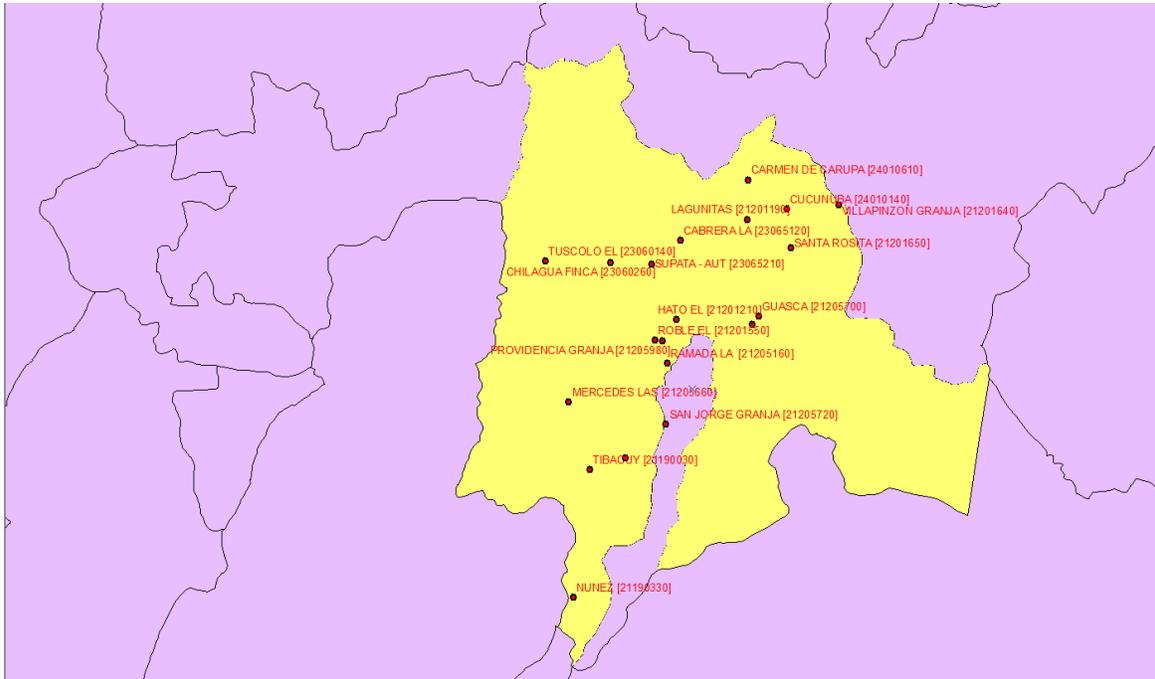


Figura 16 Estaciones seleccionadas en Cundinamarca para estudiar la variable de precipitación total anual – Fuente: Elaboración propia en ArcGIS

Para la variable de precipitación máxima en 24 horas, se obtuvo un total de 20 estaciones hidroclimáticas que se emplearon para el análisis comparativo de distribuciones estadísticas del presente trabajo de grado, se definió una caracterización de la zona de estudio, donde las altitudes de las estaciones varían desde los 810 m.s.n.m. hasta los 3160 m.s.n.m. y las precipitaciones multianuales máximas en 24 horas están en un rango de 33.7 mm/día hasta los 88.63 mm/ día. En todos los casos la distribución de la lluvia anual es de tipo bimodal, lo cual indica que al año se presentan dos picos de meses lluviosos. A continuación, se presentan las estaciones definitivas, que se emplearon en este trabajo de grado:

Tabla 15 Estaciones hidroclimatológicas definitivas para precipitación máxima en 24 horas, empleadas en el presente trabajo de grado

Código	Estación	Municipio	Altitud (m.s.n.m)	Años Registro
24015120	ISLA DEL SANTUARIO	Fúquene	2580	32
21205660	MERCEDES - AUT	Anapoima	810	48
21190330	NUNEZ	Cabrera	1950	30
24010610	CARMEN DE CARUPA	Carmen De Carupa	2970	36
23060160	SAN PABLO	Caparrapí	1200	38
24010140	CUCUNUBA	Cucunubá	2620	44
23060180	PENON EL	El Peñón	1400	45
21190310	PINAR EL	Fusagasugá	1900	30
23060140	TUSCOLO EL	Guaduas	985	41
21205700	GUASCA	Guasca	2750	45
21201610	SAN ISIDRO	Guasca	2698	30
23065120	CABRERA LA	Pacho	1920	30
23060260	CHILAGUA FINCA	Nocaima	1500	32
21205720	SAN JORGE GRANJA	Soacha	2900	49
23065210	SUPATA	Supatá	1798	40
21201210	HATO EL	Tenjo	2575	41
21201190	LAGUNITAS	Tausa	3100	33
21190030	TIBACUY	Tibacuy	1635	33
21201640	VILLAPINZON GRANJA	Villapinzón	2745	32
21201060	PANTANO REDONDO 1	Zipaquirá	3160	30

Así mismo, se presenta en la Figura 17 el mapa del departamento de Cundinamarca con las estaciones seleccionadas para estudiar la variable de precipitación máxima en 24 horas:

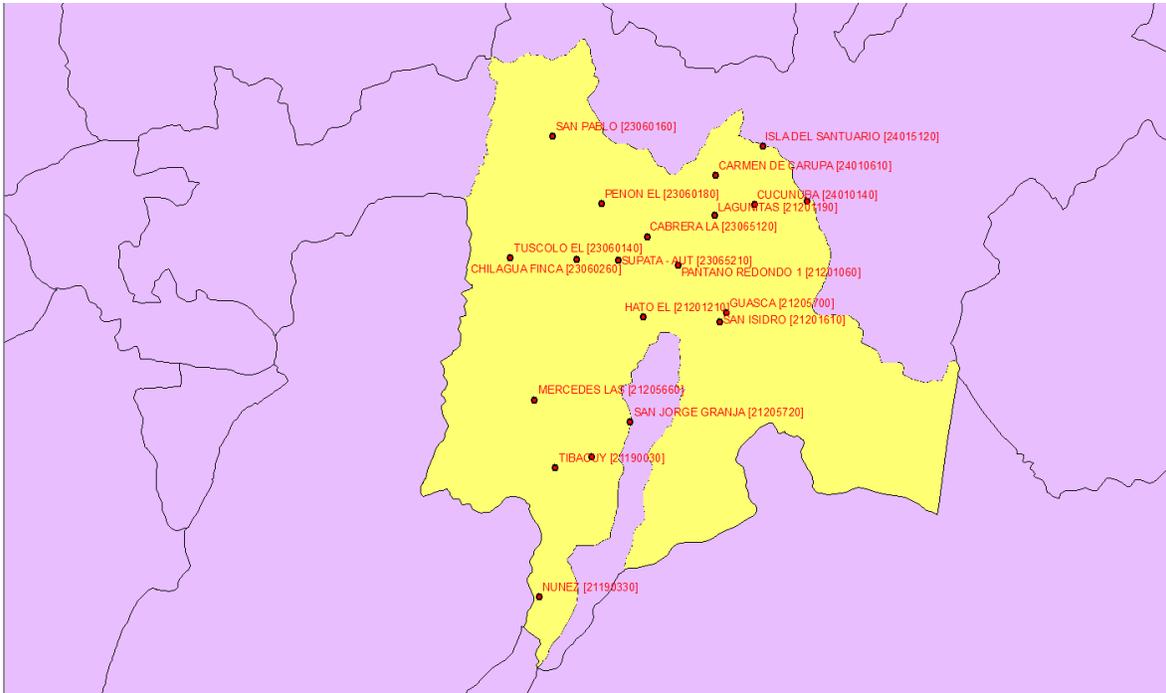


Figura 17 Estaciones seleccionadas en Cundinamarca para estudiar la variable de precipitación total anual – Fuente: Elaboración propia en ArcGIS

3.6. Procedimiento para aplicar las distribuciones estadísticas y seleccionar los mejores ajustes

Utilizando el programa EasyFit 5.6, se ajustaron las series de datos de cada una de las estaciones definidas en la sección anterior, y luego se organizaron de menor a mayor valor los estadísticos que se obtuvieron por la prueba de bondad de ajuste de Kolmogorov - Smirnov y Anderson-Darling. A continuación, en la Figura 18 se presenta la interfaz gráfica del programa EasyFit 5.6, en la cual se introducen los datos en una hoja de cálculo compuesta por casillas que tienen una localización definida por el número de fila y la letra correspondiente a la columna:

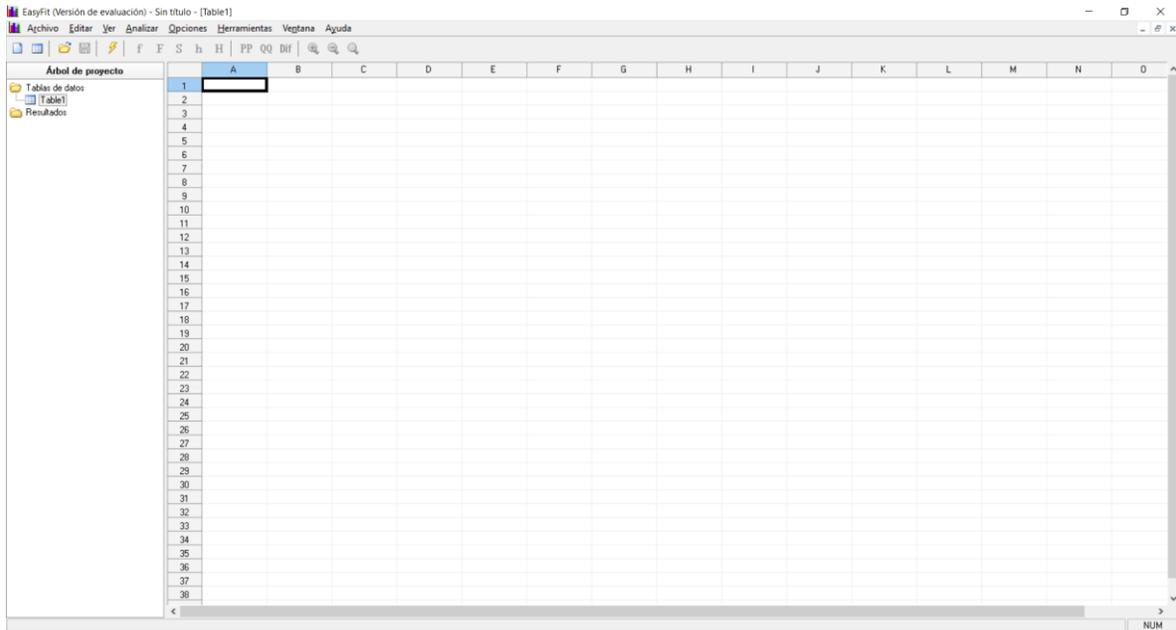


Figura 18 Interfaz gráfica del programa de distribuciones estadísticas EASYFIT 5.6

En el siguiente ejemplo se utilizaron los datos de la serie de precipitaciones totales anuales de la estación Núñez (21190330), los cuales se ubicaron en la columna A y luego se ejecutó el programa haciendo *click* en el ícono con forma de rayo localizado en la parte superior izquierda.

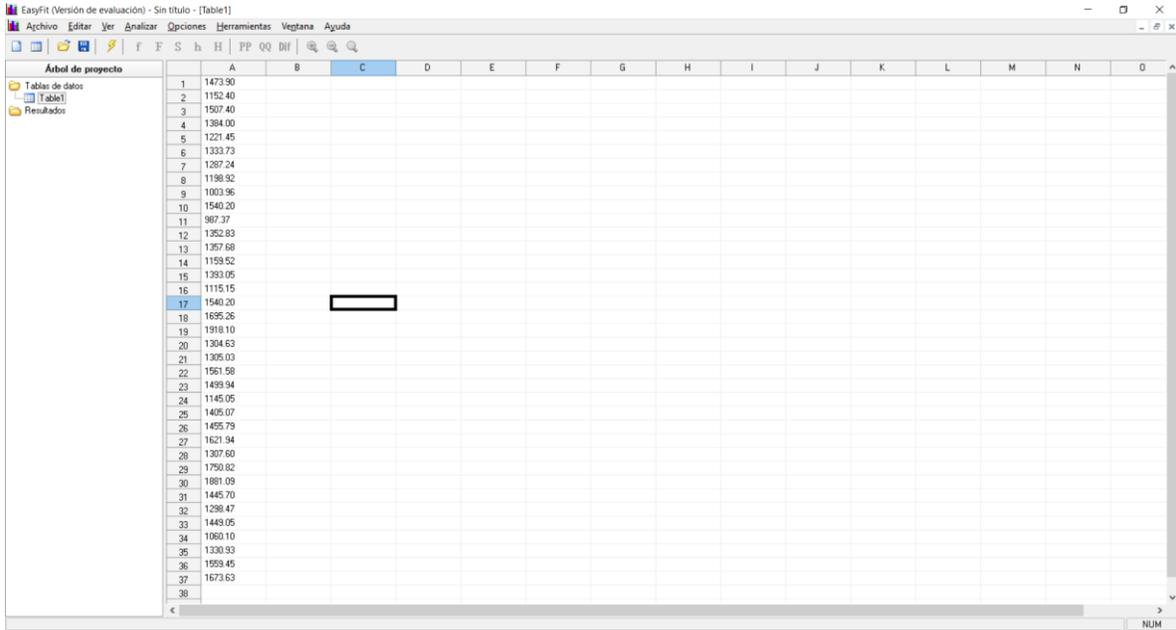


Figura 19 Serie de precipitación total anual estación Núñez ingresada en el software EasyFit (21190330)

Luego de ejecutar el programa, se muestra por orden de mejor a peor ajuste las distribuciones estadísticas analizadas, tanto por la prueba de Kolmogorov - Smirnov, como por Anderson – Darling.

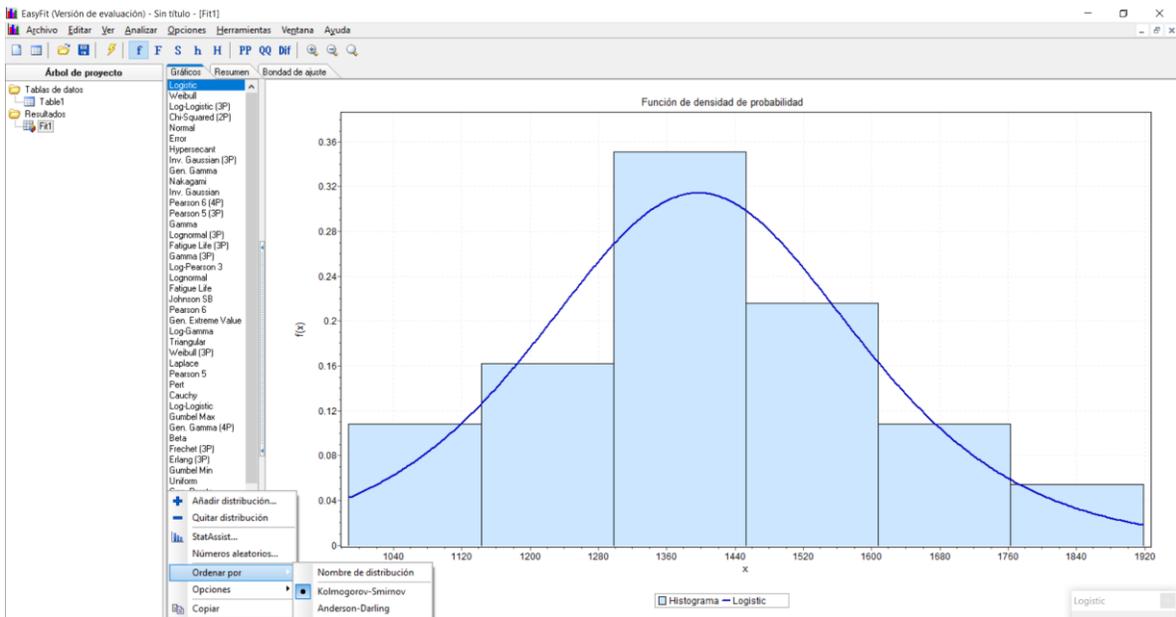


Figura 20 Distribuciones estadísticas ordenadas de mejor a peor ajuste por Kolmogorov-Smirnov para la estación Núñez (21190330)

Para la prueba de Kolmogorov – Smirnov, los mejores 5 ajustes reportados por el programa son Logistic, Weibull, Log-Logistic (3P), Chi-cuadrado (2P) y Normal.

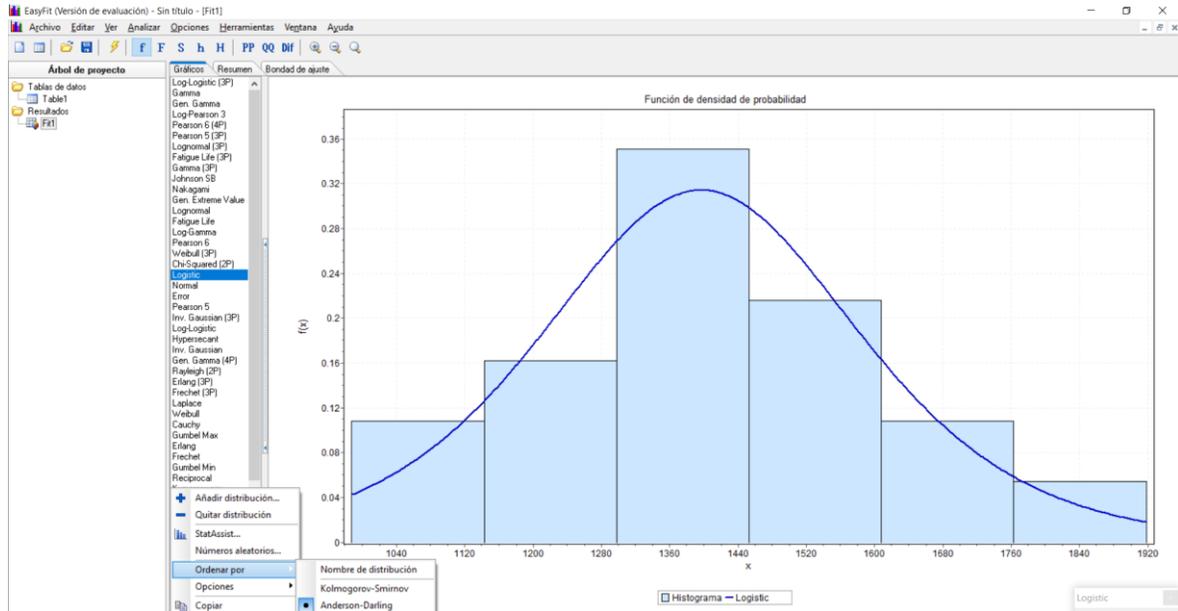


Figura 21 Distribuciones estadísticas ordenadas de mejor a peor ajuste por Anderson - Darling para la estación Núñez (21190330)

Para la prueba de Anderson - Darling, los mejores ajustes reportados por el programa son Log-Logistic (3P), Gamma, Generalizada Gamma, Log Pearson 3 y Pearson 6 (4P).

Adicionalmente para visualizar los valores de los parámetros de cada distribución estadística se debe haciendo *click* en la pestaña Resumen según lo muestra la Figura 22.

#	Distribución	Parámetros
1	Beta	$\alpha_1=1.4682$ $\alpha_2=1.8761$ $a=987.37$ $b=1918.1$
2	Burr	$k=0.00169$ $\alpha=1704.5$ $\beta=988.64$
3	Burr (4P)	$k=0.22606$ $\alpha=0.79346$ $\beta=1.3985$ $\gamma=987.37$
4	Cauchy	$\sigma=129.15$ $\mu=1380.9$
5	Chi-Squared	$v=1396$
6	Chi-Squared (2P)	$v=24272$ $\gamma=-22876.0$
7	Dagum	$k=133.97$ $\alpha=1.1748$ $\beta=10.688$
8	Dagum (4P)	$k=2.5416$ $\alpha=0.29812$ $\beta=0.43826$ $\gamma=987.37$
9	Erlang	$m=38$ $\beta=35.852$
10	Erlang (3P)	$m=23$ $\beta=45.738$ $\gamma=323.84$
11	Error	$k=2.0166$ $\sigma=223.78$ $\mu=1396.7$
12	Error Function	$h=0.00316$
13	Exponential	$\lambda=7.1597E-4$
14	Exponential (2P)	$\lambda=0.00244$ $\gamma=987.37$
15	Fatigue Life	$\alpha=0.15893$ $\beta=1379.3$
16	Fatigue Life (3P)	$\alpha=0.13329$ $\beta=1642.5$ $\gamma=-260.38$
17	Frechet	$\alpha=7.2608$ $\beta=1268.6$
18	Frechet (3P)	$\alpha=1.0850E+8$ $\beta=2.1589E+10$ $\gamma=-2.1589E+10$
19	Gamma	$\alpha=38.957$ $\beta=35.852$
20	Gamma (3P)	$\alpha=23.457$ $\beta=45.738$ $\gamma=323.84$
21	Gen. Extreme Value	$k=-0.18257$ $\sigma=212.17$ $\mu=1307.2$
22	Gen. Gamma	$k=1.0054$ $\alpha=39.74$ $\beta=35.852$

Figura 22 Valor de los parámetros de las distribuciones aplicadas a la serie de precipitaciones totales anuales de la estación Núñez (21190330)

Finalmente, al hacer click en la pestaña “Bondad de ajuste” se presentan en la columna “estadística” los estadísticos por Kolmogorov Smirnov y Anderson Darling de cada distribución estadística (Ver Figura 23). En la columna “rango” se muestra la posición en la cual está el ajuste obtenido para ambas pruebas. Es importante resaltar que 1 hace referencia a la mejor bondad de ajuste y a medida que aumenta este valor, la bondad del ajuste es peor. Para este ejemplo de la estación Núñez, el mejor ajuste obtenido mediante la prueba de Kolmogorov Smirnov es para la distribución Logistic y para la prueba de Anderson Darling fue la distribución log Logistic (3P).

#	Distribución	Kolmogorov Smirnov		Anderson Darling		Chi-cuadrado	
		Estadística	Rango	Estadística	Rango	Estadística	Rango
1	Beta	0.1131	34	2.2258	40	0.78378	32
2	Burr	0.28965	47	39.348	55	N/A	
3	Burr (4P)	0.47674	51	15.381	52	N/A	
4	Cauchy	0.10093	30	0.51444	33	0.01153	1
5	Chi-Squared	0.33353	50	54.621	56	26.622	46
6	Chi-Squared (2P)	0.06839	4	0.18204	18	0.27721	10
7	Dagum	0.5191	54	12.993	51	101.62	50
8	Dagum (4P)	0.56966	55	20.483	54	N/A	
9	Erlang	0.14185	41	0.56885	35	1.0437	35
10	Erlang (3P)	0.12156	36	0.33151	29	0.29024	12
11	Error	0.06965	6	0.18997	21	0.25628	8
12	Error Function	0.99999	59	777.25	60	N/A	
13	Exponential	0.50684	52	12.172	49	84.352	48
14	Exponential (2P)	0.27609	46	5.1455	44	11.248	42
15	Fatigue Life	0.08866	20	0.16192	14	0.34068	20
16	Fatigue Life (3P)	0.08439	16	0.15206	8	0.32054	16
17	Frechet	0.16343	43	0.94931	36	3.1126	40
18	Frechet (3P)	0.12053	35	0.34931	30	1.3966	37
19	Gamma	0.08355	14	0.14539	2	0.0522	3
20	Gamma (3P)	0.0858	17	0.15449	9	0.32458	18
21	Gen. Extreme Value	0.09036	23	0.15739	12	0.05702	5
22	Gen. Gamma	0.08041	9	0.14788	3	0.04964	2
23	Gen. Gamma (4P)	0.11229	33	0.27548	27	0.19855	6
24	Gen. Pareto	0.12699	39	7.8823	47	N/A	
25	Gumbel Max	0.10618	32	0.54298	34	0.95329	34

Figura 23 Valores de los estadísticos por cada distribución estadística y posición en el ranking

Para ver de manera regional el desempeño de cada una de las distribuciones estadísticas, se propuso realizar un sistema de puntaje ponderado donde la mejor distribución estadística de cada serie tuviera un puntaje mayor y la quinta mejor distribución tuviera el puntaje mínimo según los siguientes criterios:

1. A la distribución estadística con mejor bondad de ajuste se le asignó un puntaje de 5.
2. A la segunda distribución estadística con mejor bondad de ajuste se le asignó un puntaje de 4.
3. A la tercera distribución estadística con mejor bondad de ajuste se le asignó un puntaje de 3.
4. A la cuarta distribución estadística con mejor bondad de ajuste se le asignó un puntaje de 2.
5. A la quinta distribución estadística con mejor bondad de ajuste se le asignó un puntaje de 1.

Luego de establecer el puntaje para los cinco mejores ajustes de cada estación, se sumaron los puntajes totales por cada distribución estadística y se realizó el análisis de los resultados obtenidos, los cuales se presentan en la siguiente sección.

4. RESULTADOS AL APLICAR LAS DISTRIBUCIONES ESTADÍSTICAS

Una vez aplicadas las 61 distribuciones estadísticas a las series anuales de precipitación total y precipitación máxima en 24 horas, se presentan los resultados para cada variable hidroclimatológica y para cada prueba de bondad de ajuste.

4.1. Precipitación total anual con prueba de Kolmogorov-Smirnov

Se desarrolló un sistema de puntuaciones, en donde las cinco distribuciones estadísticas para cada serie de datos con mejor bondad de ajuste, recibieron un puntaje de 5 a 1, en donde 5 fue asignado a la que ocupó el primer puesto y 1 a la que ocupó el quinto puesto. La Tabla 16 presenta para cada una de las 20 estaciones seleccionadas con series anuales de precipitación total, el puntaje obtenido por la prueba de bondad de ajuste de Kolmogorov Smirnov.

Tabla 16 Puntajes obtenidos por las distribuciones estadísticas (61), organizados para los 5 mejores ajustes, para cada una de las estaciones (20) seleccionadas. Series anuales de precipitación total. Prueba de Kolmogorov Smirnov

Estación	Distribución	Puntaje
MERCEDES - AUT [21205660]	Burr	5
	Gen. Extreme Value	4
	Triangular	3
	Johnson SB	2
	Weibull (3P)	1
NUNEZ [21190330]	Logistic	5
	Weibull	4
	Log-Logistic (3P)	3
	Chi-Squared (2P)	2
	Normal	1
CARMEN DE CARUPA [24010610]	Log-Logistic (3P)	5
	Hypersecant	4
	Cauchy	3
	Error	2
	Laplace	1
CUCUNUBA [24010140]	Burr	5
	Log-Logistic (3P)	4
	Log-Logistic	3
	Frechet (3P)	2
	Pearson 5 (3P)	1
PINAR EL [21190310]	Gen. Gamma	5
	Erlang (3P)	4
	Pert	3
	Johnson SU	2
	Pearson 5 (3P)	1

Tabla 16. Continuación

Estación	Distribución	Puntaje
RAMADA LA [21205160]	Log-Logistic (3P)	5
	Johnson SU	4
	Logistic	3
	Error	2
	Gen. Gamma	1
TUSCOLO EL [23060140]	Log-Logistic (3P)	5
	Nakagami	4
	Gen. Gamma	3
	Burr	2
	Gamma	1
GUASCA [21205700]	Gen. Gamma (4P)	5
	Gamma	4
	Johnson SB	3
	Weibull (3P)	2
	Triangular	1
SAN ISIDRO [21201610]	Normal	5
	Chi-Squared (2P)	4
	Error	3
	Nakagami	2
	Inv. Gaussian	1
ROBLE EL [21201550]	Log-Logistic (3P)	5
	Gumbel Max	4
	Johnson SU	3
	Gen. Extreme Value	2
	Pearson 6	1
CABRERA LA [23065120]	Laplace	5
	Cauchy	4
	Hypersecant	3
	Log-Logistic (3P)	2
	Error	1
CHILAGUA FINCA [23060260]	Weibull	5
	Chi-Squared (2P)	4
	Cauchy	3
	Normal	2
	Logistic	1
SAN JORGE GRANJA [21205720]	Weibull	5
	Burr	4
	Error	3
	Chi-Squared (2P)	2
	Logistic	1

Tabla 16. Continuación

Estación	Distribución	Puntaje
SANTA ROSITA [21201650]	Gumbel Max	5
	Log-Logistic	4
	Pearson 5	3
	Weibull	2
	Log-Gamma	1
SUPATA [23065210]	Johnson SB	5
	Gen. Pareto	4
	Pearson 6 (4P)	3
	Gen. Extreme Value	2
	Error	1
HATO EL [21201210]	Burr	5
	Erlang (3P)	4
	Log-Logistic (3P)	3
	Chi-Squared (2P)	2
	Lognormal (3P)	1
LAGUNITAS [21201190]	Pearson 5 (3P)	5
	Lognormal (3P)	4
	Fatigue life (3P)	3
	Gamma (3P)	2
	Chi-Squared (2P)	1
PROVIDENCIA GRANJA [21205980]	Hypersecant	5
	Logistic	4
	Log-Logistic (3P)	3
	Error	2
	Chi-Squared (2P)	1
TIBACUY [21190030]	Cauchy	5
	Log-Logistic (3P)	4
	Log-Logistic	3
	Frechet (3P)	2
	Pearson 6 (4P)	1
VILLAPINZON GRANJA [21201640]	Log-Logistic (3P)	5
	Gumbel Max	4
	Gen. Extreme Value	3
	Johnson SU	2
	Pearson 6	1

De acuerdo con la metodología planteada en el presente trabajo de grado, el siguiente paso consistió en sumar los puntajes para cada una de las 61 distribuciones estadísticas. En la Tabla 17 se presentan los puntajes totales obtenidos para cada distribución estadística. Adicionalmente se agregó una columna con el número de veces que aparece en el *ranking* cada distribución estadística.

En color rojo se resaltan los cinco puntajes más altos y las cinco frecuencias más altas en el *ranking*. Adicionalmente se realiza el gráfico del puntaje para cada distribución estadística en la Figura 24 y el gráfico del número de apariciones en *ranking* por cada distribución estadística en la Figura 25.

Tabla 17 Puntajes totales obtenidos por las distribuciones estadísticas (61). Series anuales de precipitación total. Prueba de Kolmogorov Smirnov

Distribución	Puntaje	No. Veces
Beta	0	0
Burr	21	5
Burr (4P)	0	0
Cauchy	15	4
Chi-Squared	0	0
Chi-Squared (2P)	16	7
Dagum	0	0
Dagum (4P)	0	0
Erlang	0	0
Erlang (3P)	8	2
Error	14	7
Error function	0	0
Exponential	0	0
Exponential (2P)	0	0
Fatigue life	0	0
Fatigue life (3P)	3	1
Frechet	0	0
Frechet (3P)	4	2
Gamma	5	2
Gamma (3P)	44	1
Gen. Extreme Value	11	4
Gen. Gamma	9	3
Gen. Gamma (4P)	5	1
Gen. Pareto	4	1
Gumbel Max	13	3
Gumbel Min	0	0
Hypersecant	12	3
Inv. Gaussian	1	1
Inv. Gaussian (3P)	0	0
Johnson SU	11	4
Kumaraswamy	0	0
Laplace	6	2
Levy	0	0
Levy (2P)	0	0
Log-Gamma	1	1
Log-Logistic	10	3
Log-Logistic (3P)	44	11
Log-Pearson 3	0	0
Logistic	14	5
Lognormal	0	0

Tabla 17. Continuación

Distribución	Puntaje	No. Veces
Lognormal (3P)	5	2
Nakagami	6	2
Normal	8	3
Pareto	0	0
Pareto 2	0	0
Pearson 5	3	1
Pearson 5 (3P)	7	3
Pearson 6	2	2
Pearson 6 (4P)	4	2
Pert	3	1
Power function	0	0
Rayleigh	0	0
Rayleigh (2P)	0	0
Reciprocal	0	0
Rice	0	0
Student's	0	0
Triangular	4	2
Uniform	0	0
Weibull	16	4
Weibull (3P)	3	2
Johnson SB	10	3

De la Figura 24 se observa que la distribución Log-Logistic (3P) es la que mejor puntaje obtuvo con 44 puntos, la segunda fue Burr con un puntaje de 21, la tercera y cuarta fueron Chi-cuadrado de 2 parámetros y Weibull con 16 puntos y la quinta fue Cauchy con 15 puntos. También se observa que 34 distribuciones estadísticas tuvieron un puntaje superior a 1, lo cual representa el 55.7% de las distribuciones estudiadas, mientras que 27 distribuciones que representan el 44.3% de las distribuciones estudiadas no obtuvo ningún puntaje. Por otra parte, la distribución normal o Gaussiana, la cual es referente para valores medios obtuvo un puntaje de 8 y se ubicó en el puesto 16.

De la Figura 25 se observa que la distribución Log-Logistic (3P) es la que más veces aparece en el ranking con 11 apariciones, la segunda y tercera fueron Chi-cuadrado (2P) y Error con 7 apariciones, la cuarta y quinta fueron Burr y Logistic con 5 apariciones. También se observa que 34 distribuciones estadísticas tuvieron al menos una aparición en el *ranking*, lo cual representa el 55.7% de las distribuciones estadísticas estudiadas, mientras que 27 distribuciones que representan el 44.3% de las distribuciones estadísticas estudiadas no obtuvieron ninguna aparición en ranking. Por otra parte, la distribución normal o Gaussiana, la cual es referente para valores medios obtuvo 3 apariciones en ranking y se ubicó en el puesto 14.

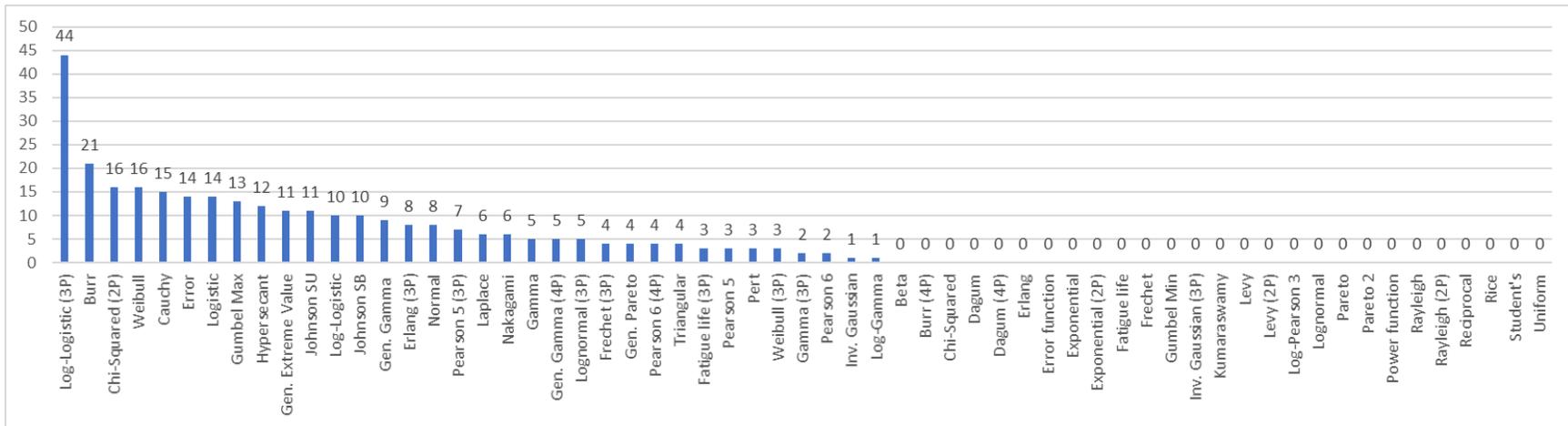


Figura 24 Puntajes totales obtenidos por las distribuciones estadísticas (61) para las series de precipitaciones totales anuales. Prueba de Kolmogorov Smirnov

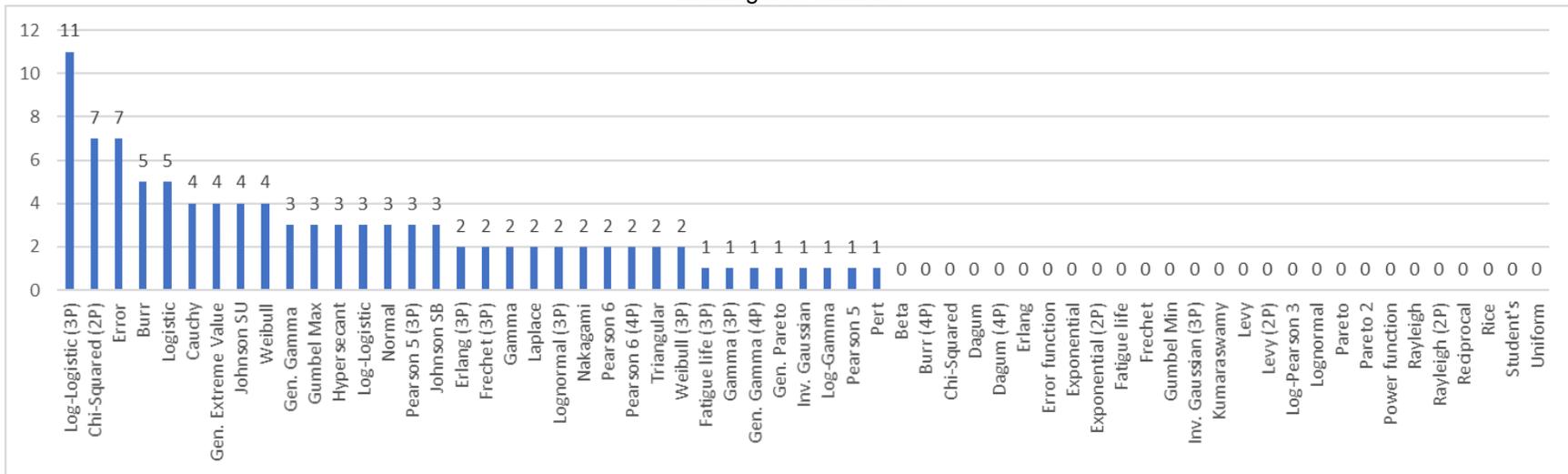


Figura 25 Número de apariciones en el ranking por las distribuciones estadísticas (61) para las series de precipitaciones totales anuales. Prueba de Kolmogorov Smirnov

En la Tabla 18 y en la Figura 26 se presentan los puntajes obtenidos para las cinco mejores distribuciones estadísticas según la prueba de bondad de ajuste por Kolmogorov-Smirnov. La línea roja punteada representa el puntaje máximo posible, el cual tiene un valor de 100 y representa el puntaje máximo que obtendría la distribución estadística ideal si llegara a ocupar el primer puesto en las 20 series de datos analizadas.

Tabla 18 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Kolmogorov-Smirnov

Distribución estadística	Puntaje
Log-Logistic (3P)	44
Burr	21
Chi-Squared (2P)	16
Weibull	16
Cauchy	15
Normal	8

Notas: Los puntajes para cada serie de datos se tomaron de 5 a 1. En este cuadro se han incluido las 5 distribuciones con mejor puntaje y la distribución normal de referencia. La distribución Log-Logistic (3P) apareció 5 veces en el primer lugar.

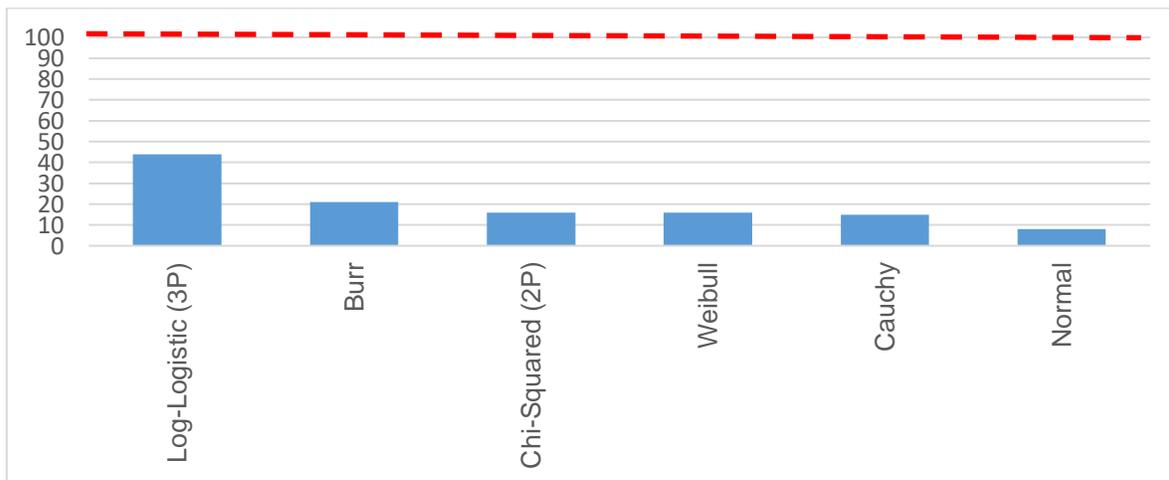


Figura 26 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Kolmogorov-Smirnov. Distribución de referencia Normal

Log-Logistic (3P) obtuvo 44 puntos de 100 máximos posibles y ocupó la primera posición, mientras que Burr ocupó la segunda posición con 21, es decir casi la mitad del puntaje con respecto a Log-Logistic (3P). Por otra parte, la distribución normal obtuvo un puntaje de 8, con lo cual se puede afirmar que Log-logistic (3P) obtuvo un desempeño 5.5 veces mayor a la distribución normal.

En la Tabla 19 y en la Figura 27 se presenta el número de apariciones para las cinco mejores distribuciones estadísticas según la prueba de bondad de ajuste por Kolmogorov-Smirnov. La línea roja punteada representa la cantidad máxima de apariciones posible, la cual tiene un valor de 20 y representa la cantidad máxima de apariciones que tendría la distribución estadística ideal si llegara a aparecer dentro de las cinco mejores distribuciones estadísticas para las 20 series de datos analizadas.

Tabla 19 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Kolmogorov-Smirnov

Distribución estadística	No. Veces en ranking
Log-Logistic (3P)	11
Error	7
Chi-Squared (2P)	7
Burr	5
Logistic	5
Normal	3

Notas: Los puntajes para cada serie de datos se tomaron de 5 a 1. En este cuadro se han incluido las 5 distribuciones con mejor puntaje y la distribución normal de referencia. La distribución Log-Logistic (3P) apareció 5 veces en el primer lugar.

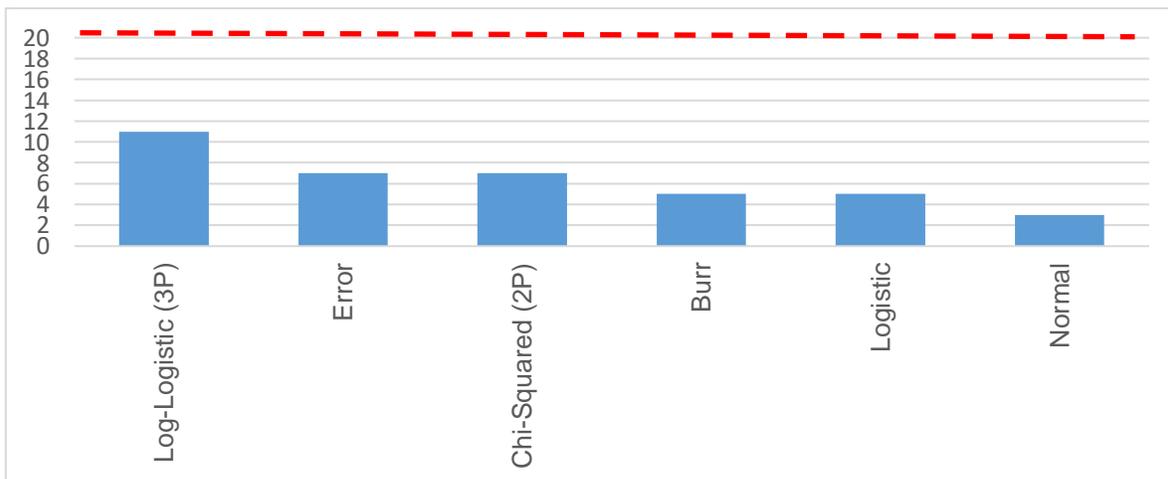


Figura 27 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Kolmogorov-Smirnov. Distribución de referencia normal

Log-Logistic (3P) obtuvo 11 apariciones en *ranking* de 20 máximas posibles, mientras que Error ocupó la segunda posición con 7 apariciones, lo que representa 0.63 veces en comparación con Log-Logistic (3P). Por otra parte, la distribución normal obtuvo 3 apariciones en *ranking*, con lo cual se puede afirmar que Log-Logistic (3P) apareció en *ranking* 3.6 veces más que la distribución normal.

Desde la Figura 28 hasta la Figura 32 se presentan los porcentajes de apariciones en el *ranking* de las distribuciones estadísticas que ocuparon los primeros, segundos, terceros, cuartos y quintos lugares respectivamente.

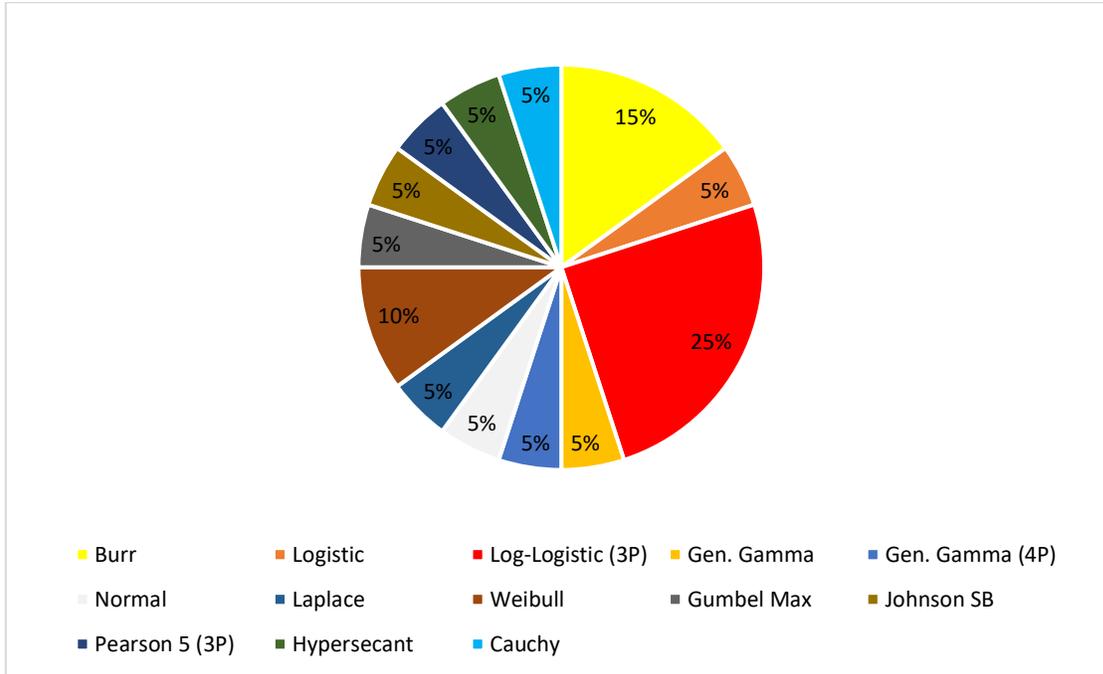


Figura 28 Distribuciones que aparecen en primera posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación total anual (20)

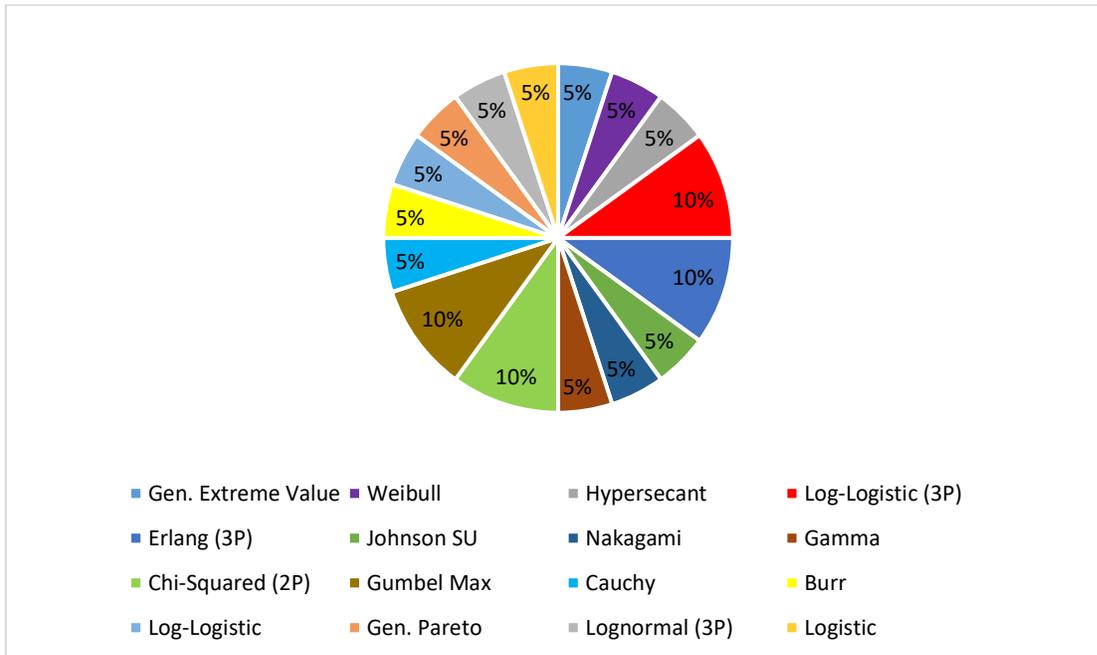


Figura 29 Distribuciones que aparecen en segunda posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación total anual (20)

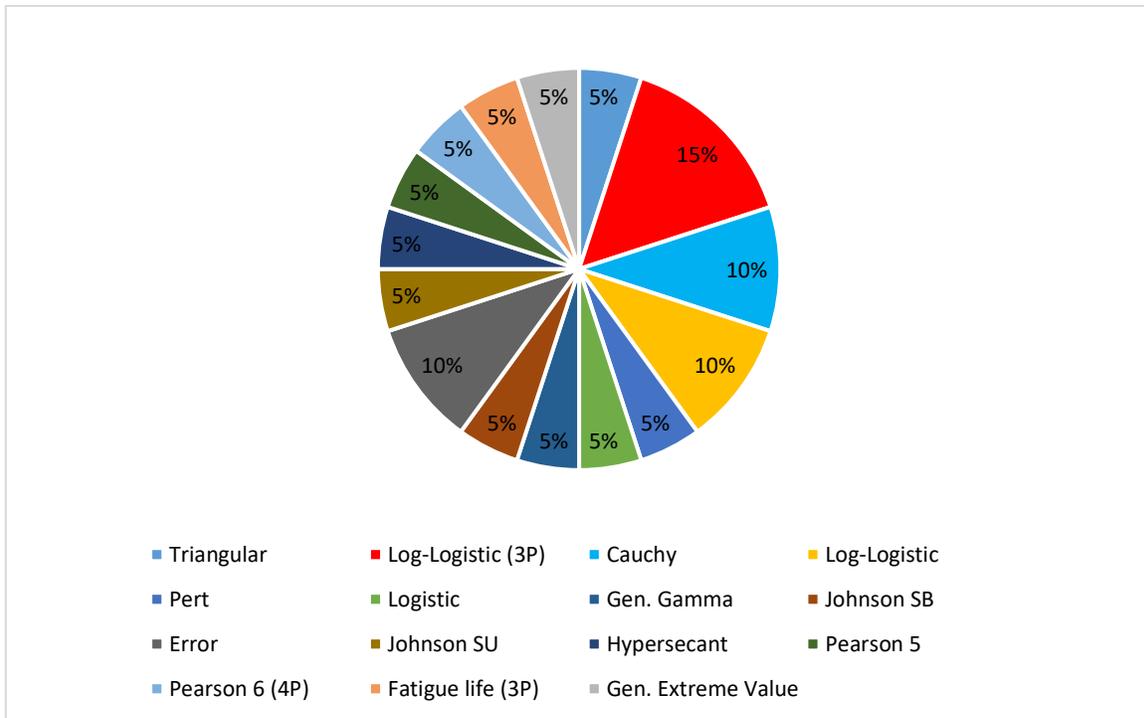


Figura 30 Distribuciones que aparecen en tercera posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación total anual (20)

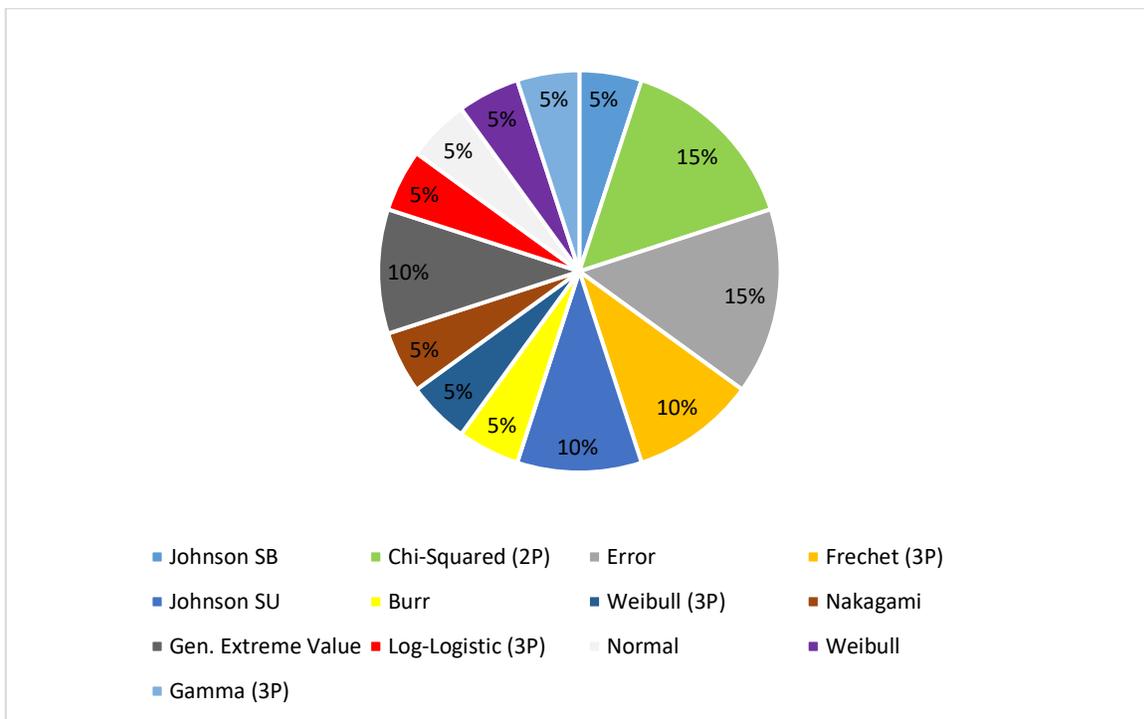


Figura 31 Distribuciones que aparecen en cuarta posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación total anual (20)

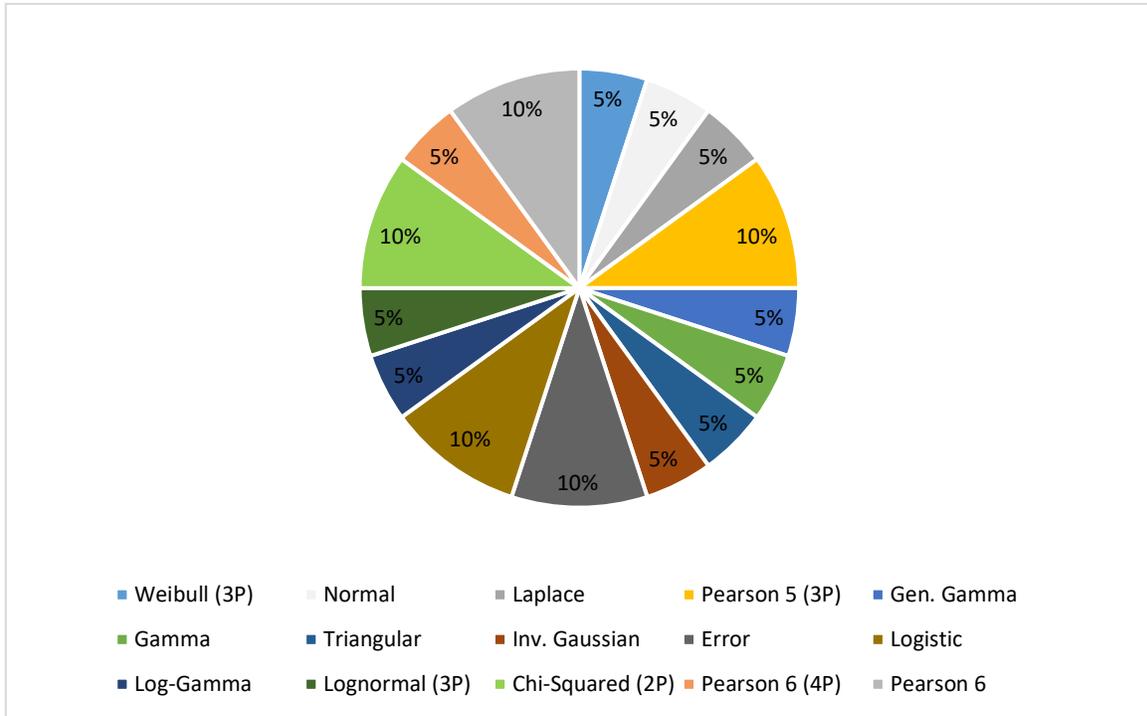


Figura 32 Distribuciones que aparecen en quinta posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación total anual (20)

En cada figura el 100% representa un valor de 20 apariciones, por lo tanto, el 5% indica que una distribución estadística solamente apareció una vez en determinada posición. La distribución Log-Logistic (3P) (identificada con color rojo) la cual apareció 11 veces en *ranking*, apareció 5 veces en la primera posición es decir el 25% del máximo posible, mientras que en segundo lugar apareció 2 veces con el 10%, en tercer lugar, apareció 3 veces con el 15% y en cuarto lugar apareció 1 vez con el 5%. En el quinto lugar no apareció ninguna vez.

La distribución Error (identificada con color magenta) la cual apareció 7 veces en *ranking*, no apareció ninguna vez en la primera y segunda posición, mientras que en tercer lugar apareció 2 veces con el 10%, en cuarto lugar, apareció 3 veces con el 15% y en quinto lugar apareció 2 veces con el 10%. A pesar de ser la segunda distribución con más número de apariciones en *ranking*, no está entre los 5 mejores puntajes debido a que las veces que apareció lo hizo en las últimas posiciones.

Finalmente, en la distribución normal (identificada con color blanco), las 3 apariciones en *ranking* se distribuyeron de la siguiente manera: 1 vez en primer lugar con el 5% del máximo posible, 1 vez en cuarto lugar con el 5% y 1 vez en el quinto lugar con el 5%.

4.2. Precipitación total anual con prueba de Anderson – Darling

Se desarrolló un sistema de puntuaciones, en donde las cinco distribuciones estadísticas para cada serie de datos con mejor bondad de ajuste, recibieron un puntaje de 5 a 1, en donde 5 fue asignado a la que ocupó el primer puesto y 1 a la que ocupó el quinto puesto. La

Tabla 20 presenta para cada una de las 20 estaciones seleccionadas con series anuales de precipitación total, el puntaje obtenido por la prueba de bondad de ajuste de Anderson-Darling.

Tabla 20 Puntajes obtenidos por las distribuciones estadísticas (61), organizados para los 5 mejores ajustes, para cada una de las estaciones (20) seleccionadas. Series anuales de precipitación total. Prueba de Anderson-Darling

Estación	Distribución	Puntaje
MERCEDES - AUT [21205660]	Burr	5
	Johnson SB	4
	Gen. Extreme Value	3
	Weibull (3P)	2
	Weibull	1
NUNEZ [21190330]	Log-Logistic (3P)	5
	Gamma	4
	Gen. Gamma	3
	Log-Pearson 3	2
	Pearson 6 (4P)	1
CARMEN DE CARUPA [24010610]	Hypersecant	5
	Log-Logistic (3P)	4
	Error	3
	Laplace	2
	Logistic	1
CUCUNUBA [24010140]	Burr	5
	Log-Logistic (3P)	4
	Log-Logistic	3
	Pearson 5	2
	Pearson 5 (3P)	1
PINAR EL [21190310]	Pearson 5 (3P)	5
	Pearson 6 (4P)	4
	Lognormal (3P)	3
	Fatigue life (3P)	2
	Gen. Gamma	1
RAMADA LA [21205160]	Hypersecant	5
	Log-Logistic (3P)	4
	Logistic	3
	Error	2
	Johnson SU	1

Tabla 20. Continuación

Estación	Distribución	Puntaje
TUSCOLO EL [23060140]	Burr	5
	Gamma	4
	Log-Logistic (3P)	3
	Log-Pearson 3	2
	Gen. Gamma	1
GUASCA [21205700]	Johnson SB	5
	Gen. Gamma (4P)	4
	Weibull (3P)	3
	Gen. Extreme Value	2
	Log-Pearson 3	1
SAN ISIDRO [21201610]	Nakagami	5
	Gen. Extreme Value	4
	Gamma	3
	Erlang	2
	Johnson SB	1
ROBLE EL [21201550]	Log-Logistic (3P)	5
	Johnson SU	4
	Pearson 5 (3P)	3
	Pearson 6 (4P)	2
	Lognormal (3P)	1
CABRERA LA [23065120]	Laplace	5
	Hypersecant	4
	Error	3
	Logistic	2
	Log-Logistic (3P)	1
CHILAGUA FINCA [23060260]	Burr	5
	Log-Logistic (3P)	4
	Hypersecant	3
	Error	2
	Pearson 6	1
SAN JORGE GRANJA [21205720]	Burr	5
	Error	4
	Chi-Squared (2P)	3
	Nakagami	2
	Normal	1
SANTA ROSITA [21201650]	Log-Pearson 3	5
	Gen. Extreme Value	4
	Gumbel Max	3
	Log-Logistic	2
	Frechet (3P)	1

Tabla 20. Continuación

Estación	Distribución	Puntaje
SUPATA [23065210]	Johnson SB	5
	Gen. Extreme Value	4
	Error	3
	Gen. Gamma (4P)	2
	Pearson 6 (4P)	1
HATO EL [21201210]	Nakagami	5
	Pearson 6 (4P)	4
	Lognormal (3P)	3
	Fatigue life (3P)	2
	Log-Pearson 3	1
LAGUNITAS [21201190]	Johnson SB	5
	Normal	4
	Error	3
	Nakagami	2
	Chi-Squared (2P)	1
PROVIDENCIA GRANJA [21205980]	Logistic	5
	Log-Logistic (3P)	4
	Hypersecant	3
	Error	2
	Johnson SU	1
TIBACUY [21190030]	Cauchy	5
	Log-Logistic (3P)	4
	Frechet (3P)	3
	Pearson 6 (4P)	2
	Pearson 5 (3P)	1
VILLAPINZON GRANJA [21201640]	Log-Logistic (3P)	5
	Johnson SU	4
	Pearson 6	3
	Pearson 5 (3P)	2
	Pearson 6 (4P)	1

De acuerdo con la metodología planteada en el presente trabajo de grado, el siguiente paso consistió en sumar los puntajes para cada una de las 61 distribuciones estadísticas. En la Tabla 21 se presentan los puntajes totales obtenidos para cada distribución estadística. Adicionalmente se agregó una columna con el número de veces que aparece en el *ranking* cada distribución estadística.

En color rojo se resaltan los cinco puntajes más altos y las frecuencias en el *ranking* más altas. Adicionalmente se realiza el gráfico del puntaje para cada distribución estadística en la Figura 33 y el gráfico del número de apariciones en ranking por cada distribución estadística en la Figura 34.

Tabla 21 Puntajes totales obtenidos por las distribuciones estadísticas (61). Series anuales de precipitación total. Prueba de Anderson-Darling

Distribución	Puntaje	No. Veces
Beta	0	0
Burr	25	5
Burr (4P)	0	0
Cauchy	5	1
Chi-Squared	0	0
Chi-Squared (2P)	4	2
Dagum	0	0
Dagum (4P)	0	0
Erlang	2	1
Erlang (3P)	0	0
Error	22	8
Error function	0	0
Exponential	0	0
Exponential (2P)	0	0
Fatigue life	0	0
Fatigue life (3P)	4	2
Frechet	0	0
Frechet (3P)	4	2
Gamma	11	3
Gamma (3P)	0	0
Gen. Extreme Value	17	5
Gen. Gamma	5	3
Gen. Gamma (4P)	6	2
Gen. Pareto	0	0
Gumbel Max	3	1
Gumbel Min	0	0
Hypersecant	20	5
Inv. Gaussian	0	0
Inv. Gaussian (3P)	0	0
Johnson SU	10	4
Kumaraswamy	0	0
Laplace	7	2
Levy	0	0
Levy (2P)	0	0
Log-Gamma	0	0
Log-Logistic	5	2
Log-Logistic (3P)	43	11
Log-Pearson 3	11	5
Logistic	11	4
Lognormal	0	0
Lognormal (3P)	7	3
Nakagami	14	4
Normal	5	2
Pareto	0	0

Tabla 21. Continuación

Distribución	Puntaje	No. Veces
Pareto 2	0	0
Pearson 5	2	1
Pearson 5 (3P)	12	5
Pearson 6	4	2
Pearson 6 (4P)	15	7
Pert	0	0
Power function	0	0
Rayleigh	0	0
Rayleigh (2P)	0	0
Reciprocal	0	0
Rice	0	0
Student's	0	0
Triangular	0	0
Uniform	0	0
Weibull	1	1
Weibull (3P)	5	2
Johnson SB	20	5

De la Figura 33 se observa que la distribución Log-Logistic (3P) es la que mejor puntaje obtuvo con 43 puntos, la segunda fue Burr con un puntaje de 25, la tercera fue Error con 22 puntos, la cuarta y quinta fueron Hypersecant y Johnson SB con 20 puntos. También se observa que 29 distribuciones estadísticas tuvieron un puntaje superior a 1, lo cual representa el 47.5% de las distribuciones estudiadas, mientras que 32 distribuciones que representan el 52.5% de las distribuciones estudiadas no obtuvo ningún puntaje. Por otra parte, la distribución normal o Gaussiana, la cual es referente para valores medios obtuvo un puntaje de 5 y se ubicó en el puesto 20.

De la Figura 34 se observa que la distribución Log-Logistic (3P) es la que más veces aparece en el ranking con 11 apariciones, la segunda fue Error con 8 apariciones, la tercera fue Pearson VI (4P) con 7 apariciones, la cuarta y quinta fueron Burr y Valor extremo generalizado con 5 apariciones. También se observa que 29 distribuciones estadísticas tuvieron al menos una aparición en el ranking, lo cual representa el 47.5% de las distribuciones estadísticas estudiadas, mientras que 32 distribuciones que representan el 52.5% de las distribuciones estadísticas estudiadas no obtuvieron ninguna aparición en ranking. Por otra parte, la distribución normal o Gaussiana, la cual es referente para valores medios obtuvo 2 apariciones en ranking y se ubicó en el puesto 22.

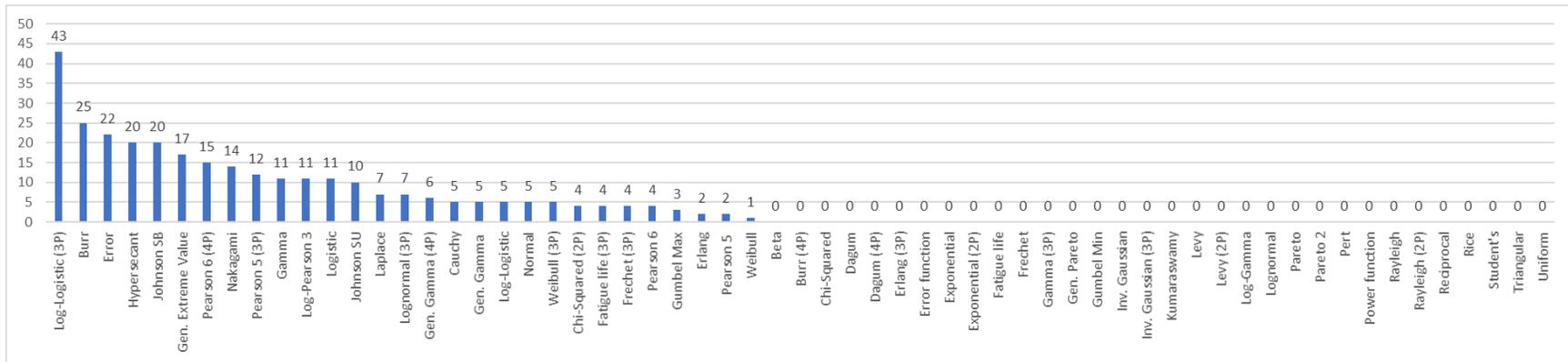


Figura 33 Puntajes totales obtenidos por las distribuciones estadísticas (61) para las series de precipitaciones totales anuales. Prueba de Anderson-Darling

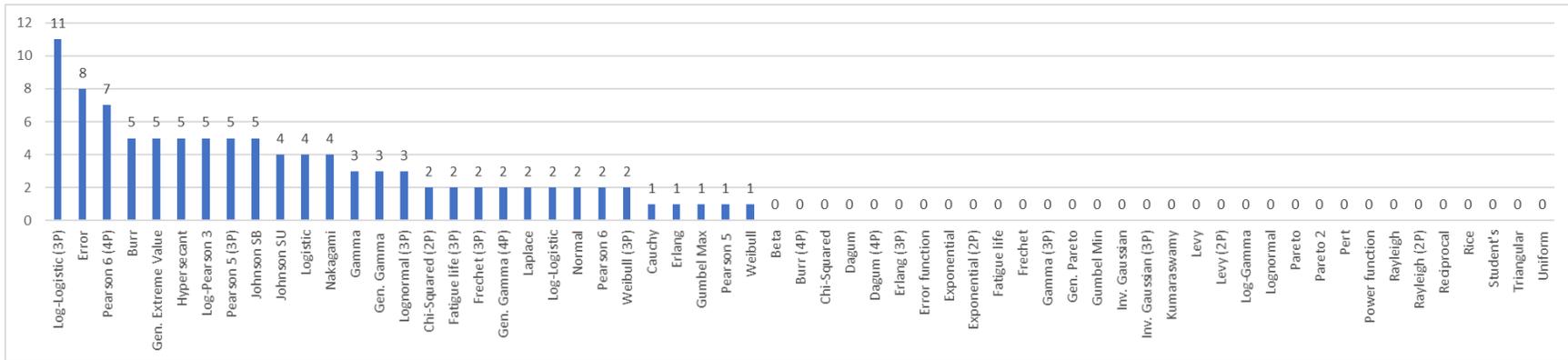


Figura 34 Número de apariciones en el ranking por las distribuciones estadísticas (61) para las series de precipitaciones totales anuales. Prueba de Anderson-Darling

En la Tabla 22 y en la Figura 35 se presentan los puntajes obtenidos para las cinco mejores distribuciones estadísticas según la prueba de bondad de ajuste por Anderson-Darling. La línea roja punteada representa el puntaje máximo posible, el cual tiene un valor de 100 y representa el puntaje máximo que obtendría la distribución estadística ideal si llegara a ocupar el primer puesto en las 20 series de datos analizadas.

Tabla 22 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Anderson-Darling

Distribución estadística	Puntaje
Log-Logistic (3P)	43
Burr	25
Error	22
Hypersecant	20
Johnson SB	20
Normal	5

Notas: Los puntajes para cada serie de datos se tomaron de 5 a 1. En este cuadro se han incluido las 5 distribuciones con mejor puntaje y la distribución normal de referencia. La distribución Log-Logistic (3P) apareció 3 veces en el primer lugar.

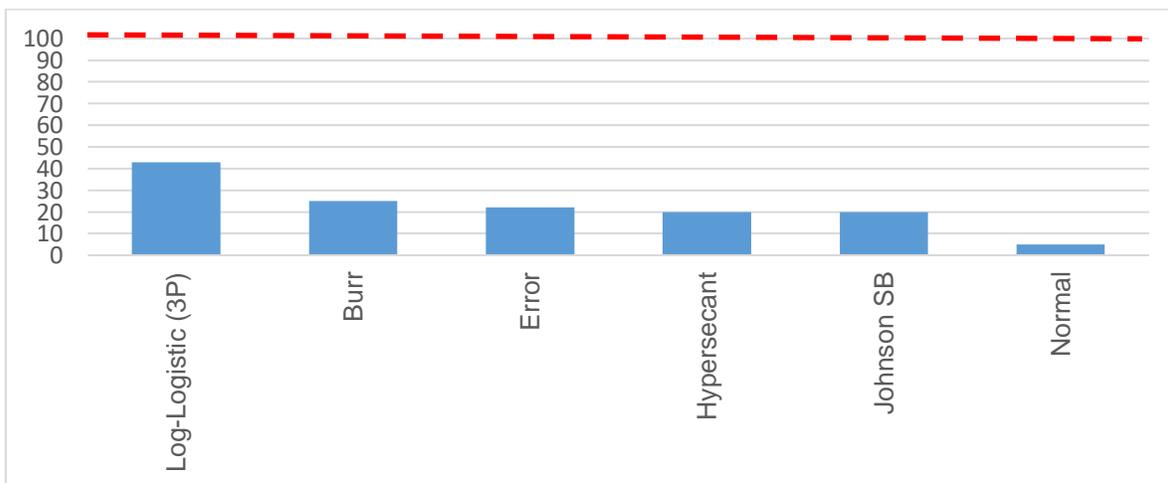


Figura 35 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Anderson-Darling. Distribución de referencia Normal.

Log-Logistic (3P) obtuvo 43 puntos de 100 máximos posibles y ocupó la primera posición, mientras que Burr ocupó la segunda posición con 25, es decir casi la mitad del puntaje con respecto a Log-Logistic (3P). Por otra parte, la distribución normal obtuvo un puntaje de 5, con lo cual se puede afirmar que Log-logistic (3P) obtuvo un desempeño 8.6 veces mayor a la distribución normal.

En la Tabla 23 y en la Figura 36 se presenta el número de apariciones para las cinco mejores distribuciones estadísticas según la prueba de bondad de ajuste por Anderson-Darling. La línea roja punteada representa la cantidad máxima de apariciones posible, la cual tiene un valor de 20 y representa la cantidad máxima de apariciones que tendría la distribución estadística ideal si llegara a aparecer dentro de las cinco mejores distribuciones estadísticas para las 20 series de datos analizadas.

Tabla 23 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Anderson-Darling

Distribución estadística	No. Veces en ranking
Log-Logistic (3P)	11
Error	8
Pearson 6 (4P)	7
Burr	5
Gen. Extreme Value	5
Normal	2

Notas: Los puntajes para cada serie de datos se tomaron de 5 a 1. En este cuadro se han incluido las 5 distribuciones con mejor puntaje y la distribución normal de referencia. La distribución Log-Logistic (3P) apareció 3 veces en el primer lugar.

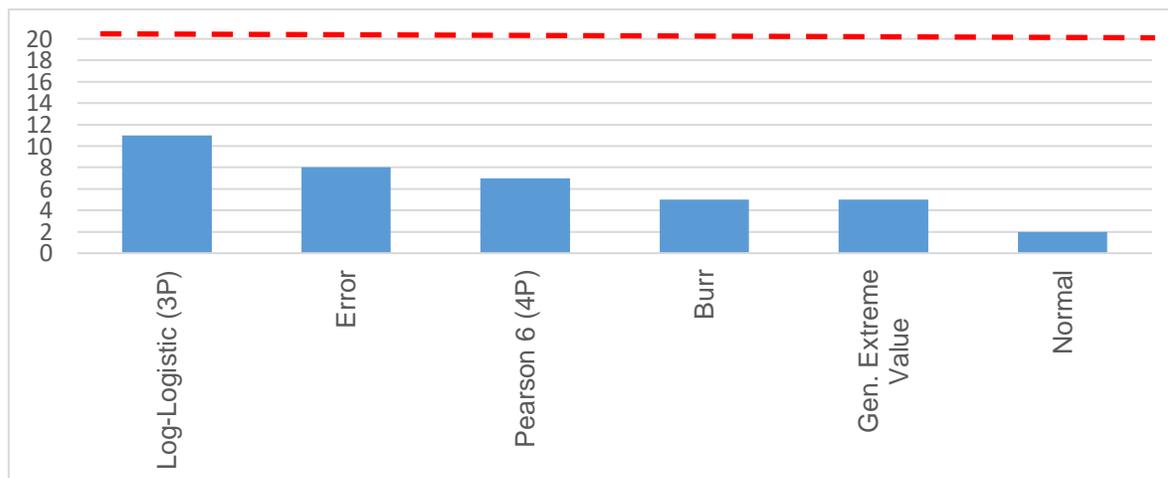


Figura 36 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones totales anuales (20). Prueba de Anderson-Darling. Distribución de referencia normal

Log-Logistic (3P) obtuvo 11 apariciones en *ranking* de 20 máximas posibles, mientras que Error ocupó la segunda posición con 8 apariciones, lo que representa 0.72 veces en comparación con Log-Logistic (3P). Por otra parte, la distribución normal obtuvo 2 apariciones en *ranking*, con lo cual se puede afirmar que Log-Logistic (3P) apareció en ranking 5.5 veces más que la distribución normal.

Desde la Figura 37 hasta la Figura 41 se presentan los porcentajes de apariciones en el *ranking* de las distribuciones estadísticas que ocuparon los primeros, segundos, terceros, cuartos y quintos lugares respectivamente.

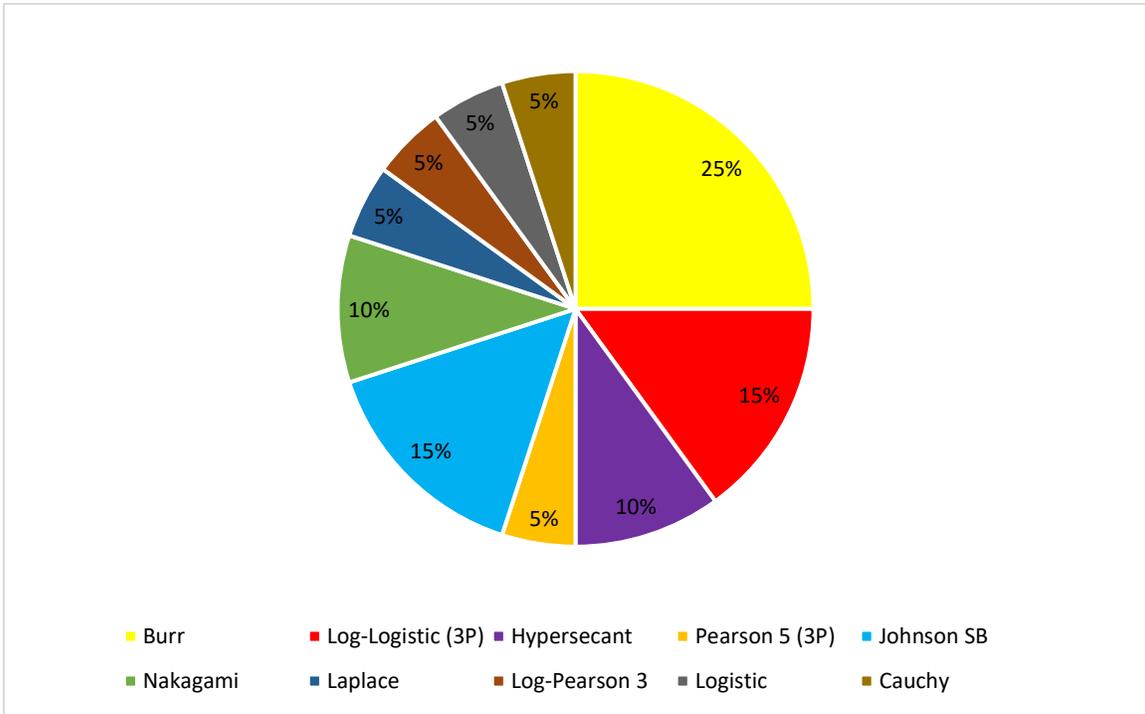


Figura 37 Distribuciones que aparecen en primera posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación total anual (20)

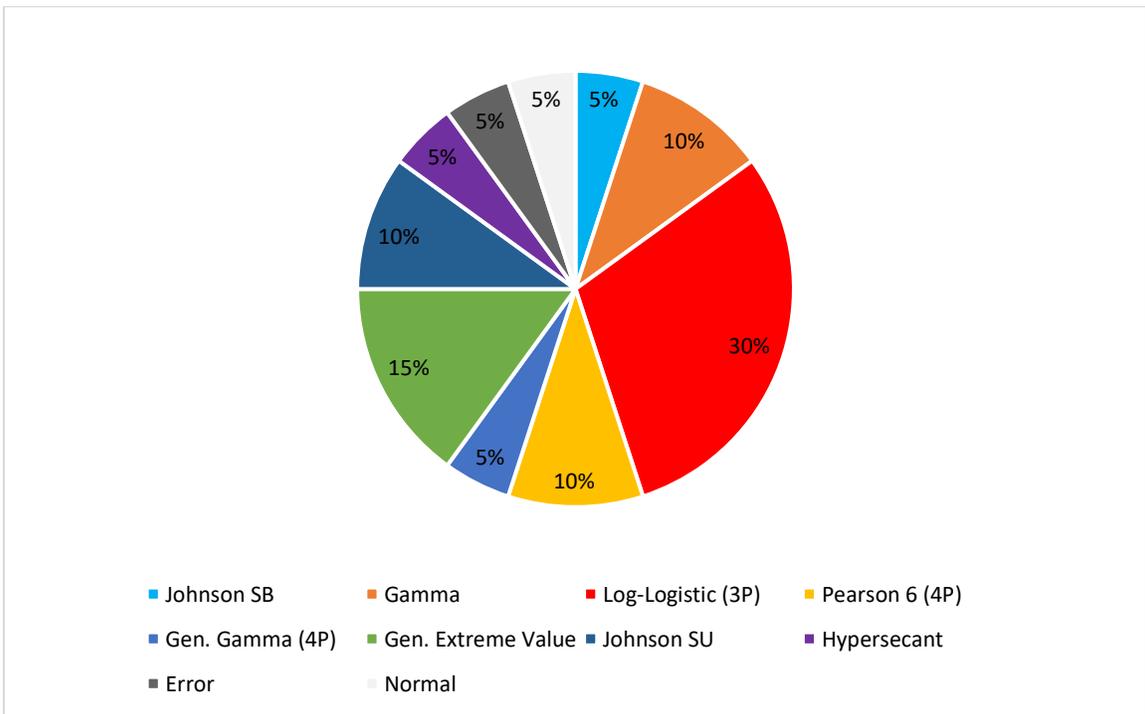


Figura 38 Distribuciones que aparecen en segunda posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación total anual (20)

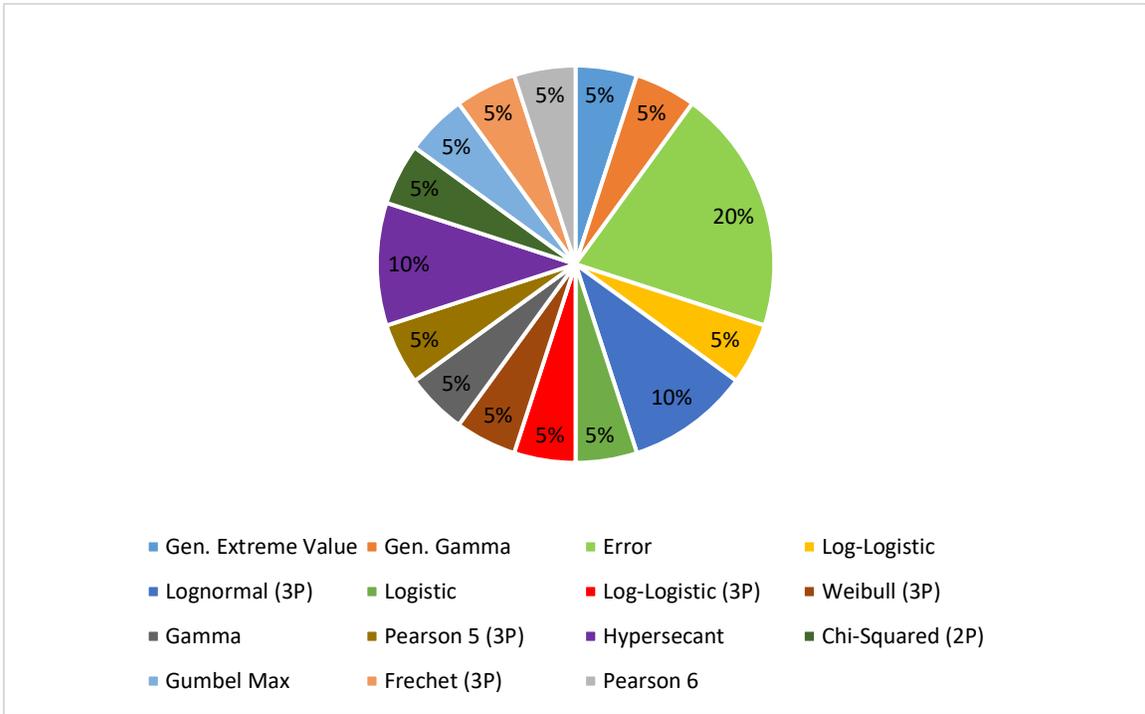


Figura 39 Distribuciones que aparecen en tercera posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación total anual (20)

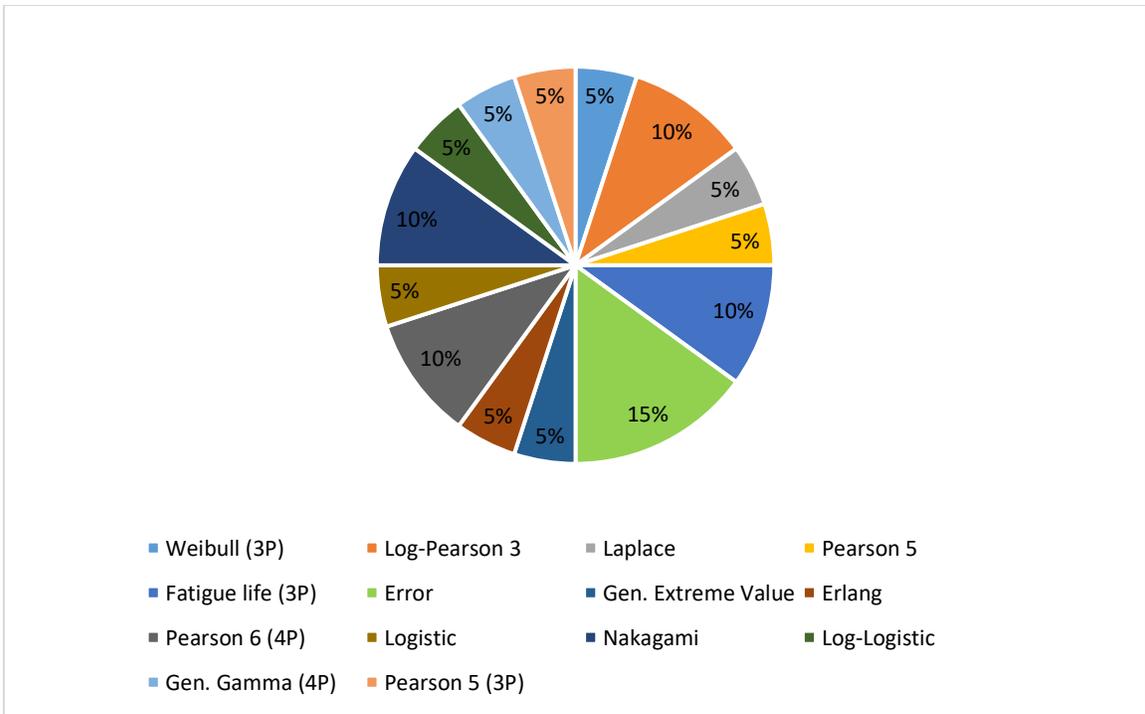


Figura 40 Distribuciones que aparecen en cuarta posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación total anual (20)

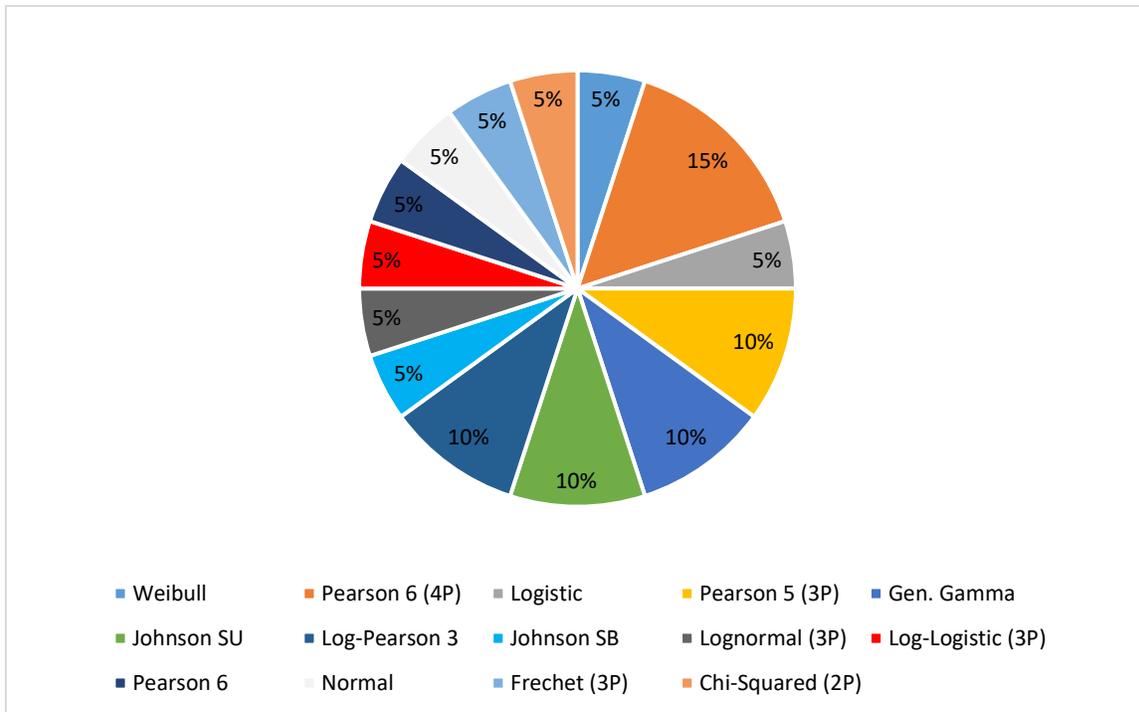


Figura 41 Distribuciones que aparecen en quinta posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación total anual (20).

En cada figura el 100% representa un valor de 20 apariciones, por lo tanto, el 5% indica que una distribución estadística solamente apareció una vez en determinada posición. La distribución Log-Logistic (3P) (identificada con color rojo) la cual apareció 11 veces en *ranking*, apareció 3 veces en la primera posición es decir el 15% del máximo posible, mientras que en segundo lugar apareció 6 veces con el 30%, en tercer lugar, apareció 1 vez con el 5% y en cuarto lugar no apareció ninguna vez y en el quinto lugar apareció una vez con el 5%.

La distribución Error (identificada con color verde claro) la cual apareció 8 veces en *ranking*, no apareció ninguna vez en la primera posición, mientras que en la segunda posición apareció 1 vez, en tercer lugar, apareció 4 veces con el 20%, en cuarto lugar, apareció 3 veces con el 15% y en quinto no apareció ninguna vez. A pesar de ser la segunda distribución con más número de apariciones en *ranking*, es la tercera con mejor puntaje debido al puntaje obtenido por aparecer en terceros y cuartos lugares.

Finalmente, en la distribución normal (identificada con color blanco), las 2 apariciones en *ranking* se distribuyeron de la siguiente manera: 1 vez en segundo lugar con el 5% del máximo posible y 1 vez en el quinto lugar con el 5%.

4.3. Precipitación máxima en 24 horas con prueba de Kolmogorov-Smirnov

Se desarrolló un sistema de puntuaciones, en donde las cinco distribuciones estadísticas para cada serie de datos con mejor bondad de ajuste, recibieron un puntaje de 5 a 1, en donde 5 fue asignado a la que ocupó el primer puesto y 1 a la que ocupó el quinto puesto. La Tabla 24 presenta para cada una de las 20 estaciones seleccionadas con series anuales de precipitación máxima en 24 horas, el puntaje obtenido por la prueba de bondad de ajuste de Kolmogorov Smirnov.

Tabla 24 Puntajes obtenidos por las distribuciones estadísticas (61), organizados para los 5 mejores ajustes, para cada una de las estaciones (20) seleccionadas. Series anuales de precipitación máxima en 24 horas. Prueba de Kolmogorov-Smirnov

Estación	Distribución	Puntaje
ISLA DEL SANTUARIO [24015120]	Burr	5
	Gumbel Max	4
	Dagum	3
	Log-Logistic (3P)	2
	Pearson 5	1
MERCEDES - AUT [21205660]	Kumaraswamy	5
	Beta	4
	Gen. Gamma (4P)	3
	Pearson 6	2
	Gen. Pareto	1
NUNEZ [21190330]	Nakagami	5
	Error	4
	Logistic	3
	Normal	2
	Burr	1
CARMEN DE CARUPA [24010610]	Lognormal (3P)	5
	Gen. Extreme Value	4
	Pearson 5 (3P)	3
	Pearson 6 (4P)	2
	Inv. Gaussian (3P)	1
SAN PABLO [23060160]	Burr	5
	Log-Logistic (3P)	4
	Gen. Extreme Value	3
	Johnson SB	2
	Frechet (3P)	1
CUCUNUBA [24010140]	Gen. Pareto	5
	Weibull (3P)	4
	Frechet	3
	Johnson SB	2
	Gen. Extreme Value	1
PENON EL [23060180]	Gen. Gamma (4P)	5
	Johnson SB	4
	Gen. Pareto	3
	Weibull (3P)	2
	Log-Pearson 3	1

Tabla 24. Continuación

Estación	Distribución	Puntaje
PINAR EL [21190310]	Fatigue life (3P)	5
	Inv. Gaussian (3P)	4
	Log-Pearson 3	3
	Lognormal (3P)	2
	Gamma (3P)	1
TUSCOLO EL [23060140]	Dagum (4P)	5
	Pearson 5 (3P)	4
	Lognormal	3
	Pearson 6	2
	Gen. Gamma	1
GUASCA [21205700]	Johnson SB	5
	Gen. Extreme Value	4
	Weibull (3P)	3
	Gen. Pareto	2
	Pert	1
SAN ISIDRO [21201610]	Gen. Pareto	5
	Nakagami	4
	Kumaraswamy	3
	Rice	2
	Johnson SB	1
CABRERA LA [23065120]	Pearson 5	5
	Pearson 6	4
	Gen. Extreme Value	3
	Weibull	2
	Lognormal	1
CHILAGUA FINCA [23060260]	Normal	5
	Weibull	4
	Gen. Extreme Value	3
	Gumbel Min	2
	Gen. Gamma	1
SAN JORGE GRANJA [21205720]	Johnson SB	5
	Gen. Extreme Value	4
	Log-Pearson 3	3
	Inv. Gaussian	2
	Burr (4P)	1
SUPATA [23065210]	Cauchy	5
	Weibull	4
	Log-Gamma	3
	Chi-Squared	2
	Lognormal	1

Tabla 24. Continuación

Estación	Distribución	Puntaje
HATO EL [21201210]	Dagum	5
	Log-Logistic (3P)	4
	Frechet (3P)	3
	Burr	2
	Pearson 5 (3P)	1
LAGUNITAS [21201190]	Weibull (3P)	5
	Pert	4
	Johnson SB	3
	Gen. Pareto	2
	Erlang (3P)	1
TIBACUY [21190030]	Gen. Gamma	5
	Gamma	4
	Pearson 5 (3P)	3
	Pearson 6	2
	Gamma (3P)	1
VILLAPINZON GRANJA [21201640]	Pearson 6 (4P)	5
	Lognormal (3P)	4
	Fatigue life (3P)	3
	Gamma (3P)	2
	Pearson 6	1
PANTANO REDONDO 1 [21201060]	Error	5
	Burr	4
	Chi-Squared (2P)	3
	Normal	2
	Dagum	1

De acuerdo con la metodología planteada en el presente trabajo de grado, el siguiente paso consistió en sumar los puntajes para cada una de las 61 distribuciones estadísticas. En la Tabla 25 se presentan los puntajes totales obtenidos para cada distribución estadística. Adicionalmente se agregó una columna con el número de veces que aparece en el *ranking* cada distribución estadística.

En color rojo se resaltan los cinco puntajes más altos y las frecuencias en el ranking más altas. Adicionalmente se realiza el gráfico del puntaje para cada distribución estadística en la Figura 42 y el gráfico del número de apariciones en ranking por cada distribución estadística en la Figura 43.

Tabla 25 Puntajes totales obtenidos por las distribuciones estadísticas (61). Series anuales de precipitación máxima en 24 horas. Prueba de Kolmogorov-Smirnov

Distribución	Puntaje	No. Veces
Beta	4	1
Burr	17	5
Burr (4P)	1	1
Cauchy	5	1
Chi-Squared	2	1
Chi-Squared (2P)	3	1
Dagum	9	3
Dagum (4P)	5	1
Erlang	0	0
Erlang (3P)	1	1
Error	9	2
Error function	0	0
Exponential	0	0
Exponential (2P)	0	0
Fatigue life	0	0
Fatigue life (3P)	8	2
Frechet	3	1
Frechet (3P)	4	2
Gamma	4	1
Gamma (3P)	6	3
Gen. Extreme Value	22	7
Gen. Gamma	7	3
Gen. Gamma (4P)	8	2
Gen. Pareto	18	6
Gumbel Max	4	1
Gumbel Min	2	1
Hypersecant	0	0
Inv. Gaussian	2	1
Inv. Gaussian (3P)	5	2
Johnson SU	0	0
Kumaraswamy	8	2
Laplace	0	0
Levy	0	0
Levy (2P)	0	0
Log-Gamma	3	1
Log-Logistic	0	0
Log-Logistic (3P)	10	3
Log-Pearson 3	7	3
Logistic	3	1
Lognormal	5	3
Lognormal (3P)	11	3
Nakagami	9	2
Normal	9	3
Pareto	0	0

Tabla 25. Continuación

Distribución	Puntaje	No. Veces
Pareto 2	0	0
Pearson 5	6	2
Pearson 5 (3P)	11	4
Pearson 6	11	5
Pearson 6 (4P)	7	2
Pert	5	2
Power function	0	0
Rayleigh	0	0
Rayleigh (2P)	0	0
Reciprocal	0	0
Rice	2	1
Student's	0	0
Triangular	0	0
Uniform	0	0
Weibull	10	3
Weibull (3P)	14	4
Johnson SB	22	7

De la Figura 42 se observa que las distribuciones valor extremo generalizada y Johnson SB son las que mejor puntaje obtuvieron con 22 puntos cada una, la tercera fue Pareto Generalizada con 18 puntos, la cuarta fue Burr con 17 puntos y la quinta fue Weibull (3P) con 14 puntos. También se observa que 41 distribuciones estadísticas tuvieron un puntaje superior a 1, lo cual representa el 67.2% de las distribuciones estudiadas, mientras que 20 distribuciones que representan el 32.8% de las distribuciones estudiadas no obtuvo ningún puntaje. Por otra parte, las distribuciones Log-Pearson 3 y Gumbel Max, las cuales son referente para ajustes de valores extremos, obtuvieron puntajes de 7 y 4 respectivamente.

De la Figura 43 se observa que las distribuciones valor extremo generalizada y Johnson SB son las que más veces aparecen en el ranking con 7 apariciones, la tercera fue Pareto generalizada con 6 apariciones, la cuarta y quinta fueron Burr y Pearson 6 con 5 apariciones. También se observa que 41 distribuciones estadísticas tuvieron al menos una aparición en el ranking, lo cual representa el 67.2% de las distribuciones estadísticas estudiadas, mientras que 20 distribuciones que representan el 32.8% de las distribuciones estadísticas estudiadas no tuvieron ninguna aparición en *ranking*. Por otra parte, las distribuciones Valor extremo generalizado, Gumbel y Log-Pearson 3 las cuales son referente para valores extremos, obtuvieron 7,1 y 3 apariciones en el *ranking* ubicándose en los puestos 1, 36 y 12 respectivamente.

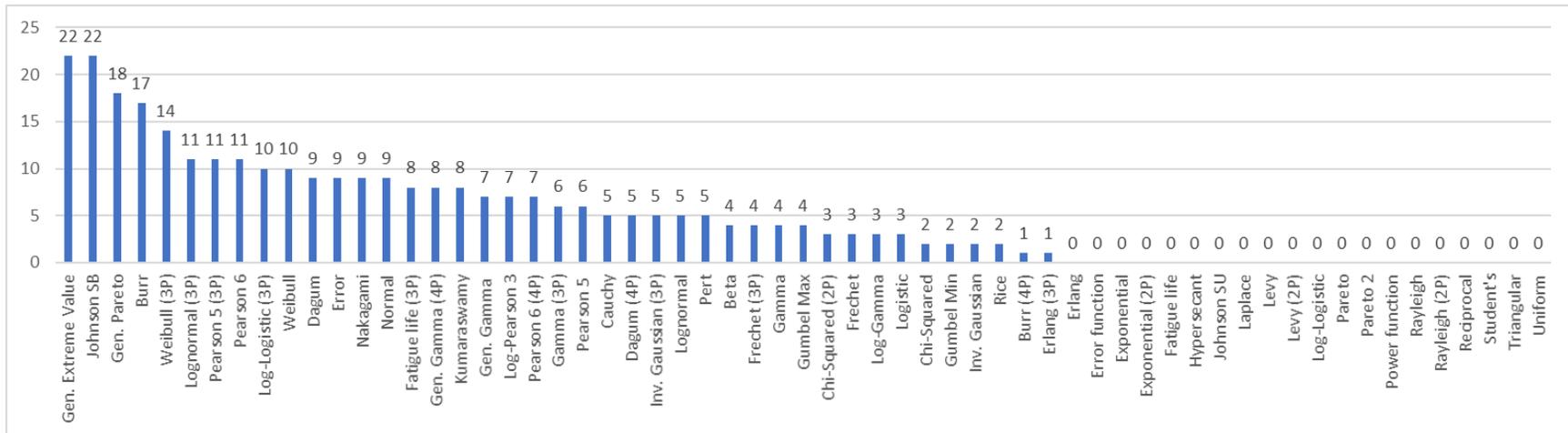


Figura 42 Puntajes totales obtenidos por las distribuciones estadísticas (61) para las series de precipitación máxima en 24 horas. Prueba de Kolmogorov-Smirnov

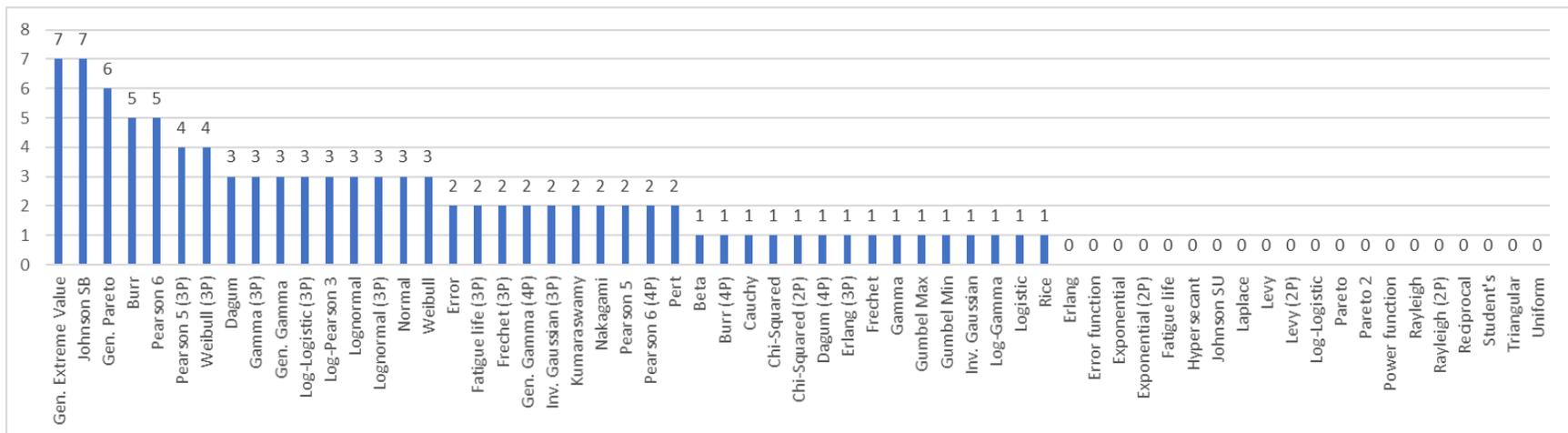


Figura 43 Número de apariciones en el ranking por las distribuciones estadísticas (61) para las series de precipitación máxima en 24 horas. Prueba de Kolmogorov-Smirnov

En la Tabla 26 y en la Figura 44 se presentan los puntajes obtenidos para las cinco mejores distribuciones estadísticas según la prueba de bondad de ajuste por Kolmogorov-Smirnov. La línea roja punteada representa el puntaje máximo posible, el cual tiene un valor de 100 y representa el puntaje máximo que obtendría la distribución estadística ideal si llegara a ocupar el primer puesto en las 20 series de datos analizadas.

Tabla 26 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Kolmogorov-Smirnov

Distribución estadística	Puntaje
Gen. Extreme Value	22
Johnson SB	22
Gen. Pareto	18
Burr	17
Weibull (3P)	14
Log-Pearson 3	7
Gumbel Max	4

Notas: Los puntajes para cada serie de datos se tomaron de 5 a 1. En este cuadro se han incluido las 5 distribuciones con mejor puntaje y las distribuciones Log-Pearson 3 y Gumbel de referencia. La distribución Johnson SB apareció 2 veces en el primer lugar, y la distribución GEV ninguna.

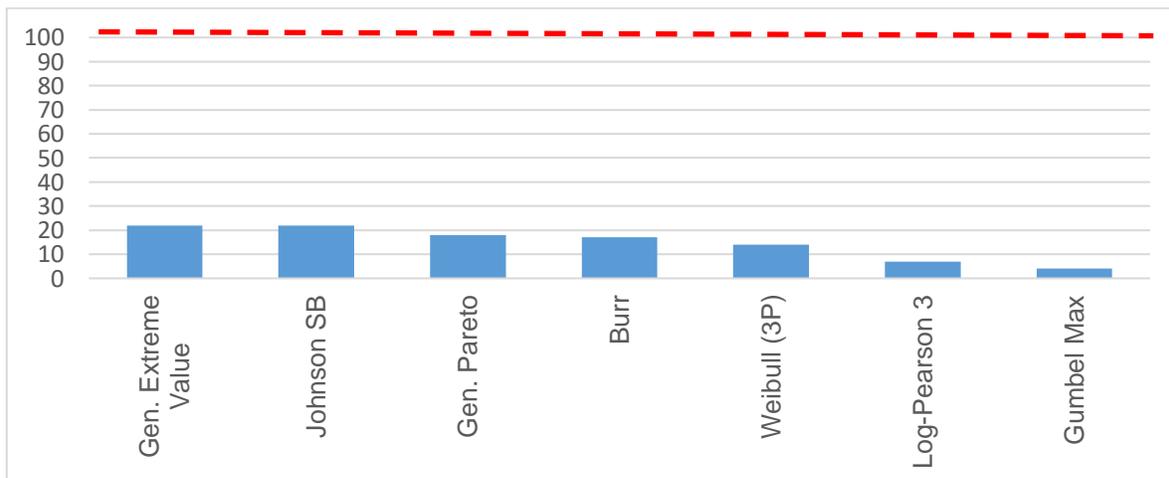


Figura 44 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Kolmogorov-Smirnov. Distribuciones Gumbel y Log Pearson 3 de referencia

Valor extremo generalizado y Johnson SB obtuvieron 22 puntos de 100 máximos posibles cada una, lo cual las ubica en el puesto 1 con el mayor puntaje obtenido. El tercer puesto fue para la distribución Pareto Generalizada la cual obtuvo 18 puntos, es decir 0.81 veces el puntaje en comparación con el primer puesto. Por otra parte, de las distribuciones que son referente para ajustar valores extremos, Log-Pearson 3 obtuvo 7 puntos, es decir casi 3 veces menos puntaje en comparación con los primeros puestos y Gumbel obtuvo 4 puntos, es decir 5.5 veces menos puntaje en comparación con los primeros puestos.

En la Tabla 27 y en la Figura 45 se presenta el número de apariciones para las cinco mejores distribuciones estadísticas según la prueba de bondad de ajuste por Kolmogorov-Smirnov. La línea roja punteada representa la cantidad máxima de apariciones posible, la cual tiene un valor de 20 y representa la cantidad máxima de apariciones que tendría la distribución estadística ideal si llegara a aparecer dentro de las cinco mejores distribuciones estadísticas para las 20 series de datos analizadas.

Tabla 27 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Kolmogorov-Smirnov

Distribución estadística	No. Veces en ranking
Gen. Extreme Value	7
Johnson SB	7
Gen. Pareto	6
Burr	5
Pearson 6	5
Log-Pearson 3	3
Gumbel Max	1

Notas: Los puntajes para cada serie de datos se tomaron de 5 a 1. En este cuadro se han incluido las 5 distribuciones con mejor puntaje y las distribuciones Log-Pearson 3 y Gumbel de referencia. La distribución Johnson SB apareció 2 veces en el primer lugar, y la distribución GEV ninguna.

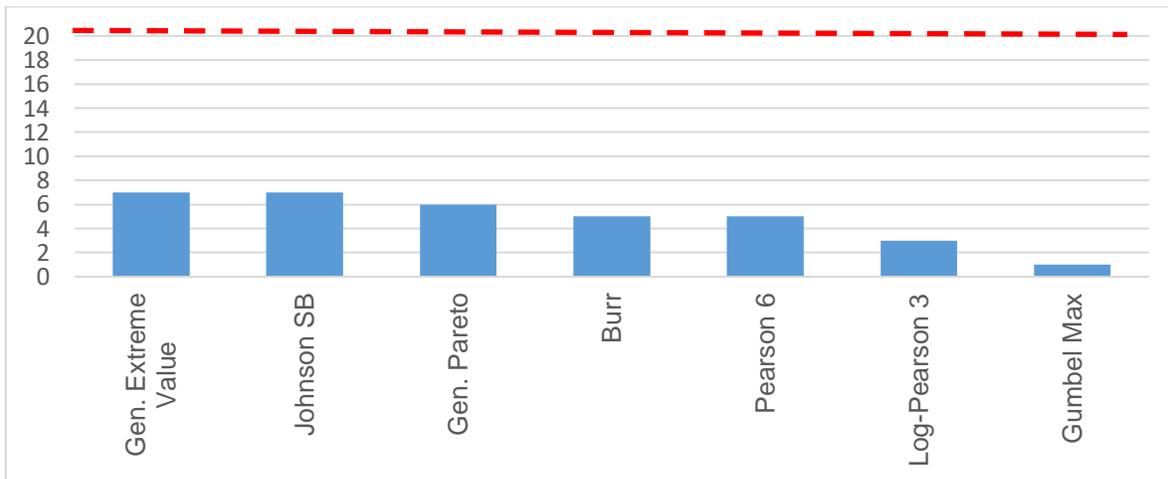


Figura 45 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Kolmogorov-Smirnov. Distribuciones Gumbel y Log Pearson 3 de referencia

Valor extremo generalizado y Johnson SB obtuvieron 7 apariciones de 20 máximas posibles cada una, este resultado coincide con los puntajes más altos y las ubica en primera posición. Por otra parte, la distribución Log-Pearson 3 obtuvo 3 apariciones, lo que equivale a 0.42 veces con respecto a las que ocuparon el primer puesto y Gumbel Max obtuvo una aparición, es decir 7 veces menos en comparación con las que ocuparon el primer lugar.

Desde la Figura 46 hasta la Figura 50 se presentan los porcentajes de apariciones en el *ranking* de las distribuciones estadísticas que ocuparon los primeros, segundos, terceros, cuartos y quintos lugares respectivamente.

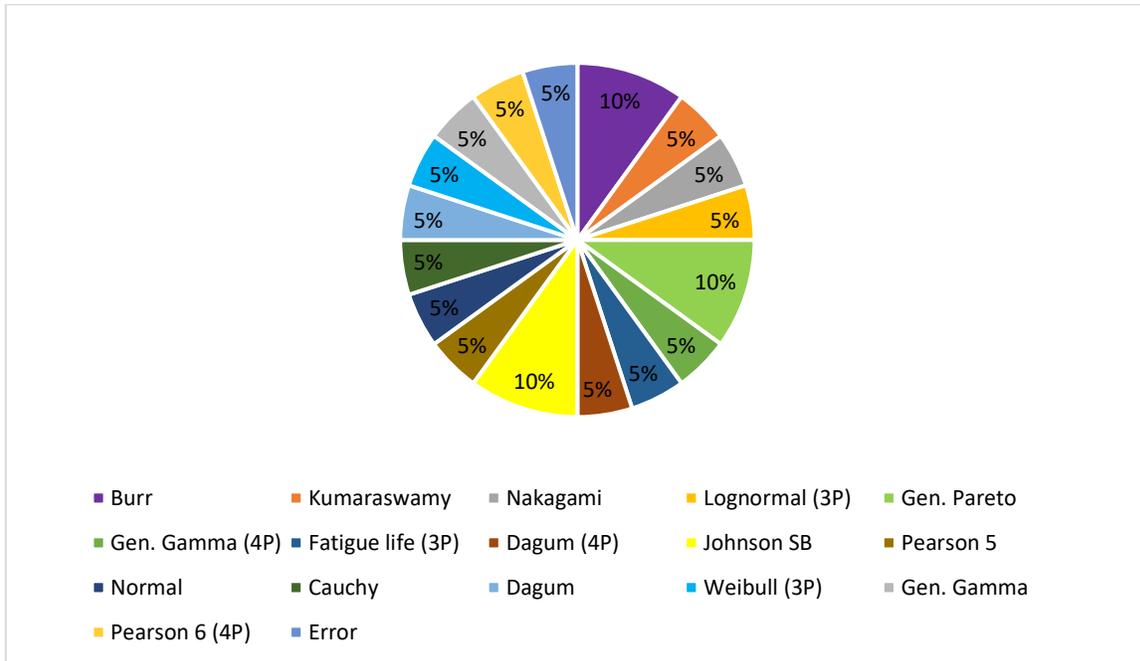


Figura 46 Distribuciones que aparecen en primera posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación máxima en 24 horas (20).

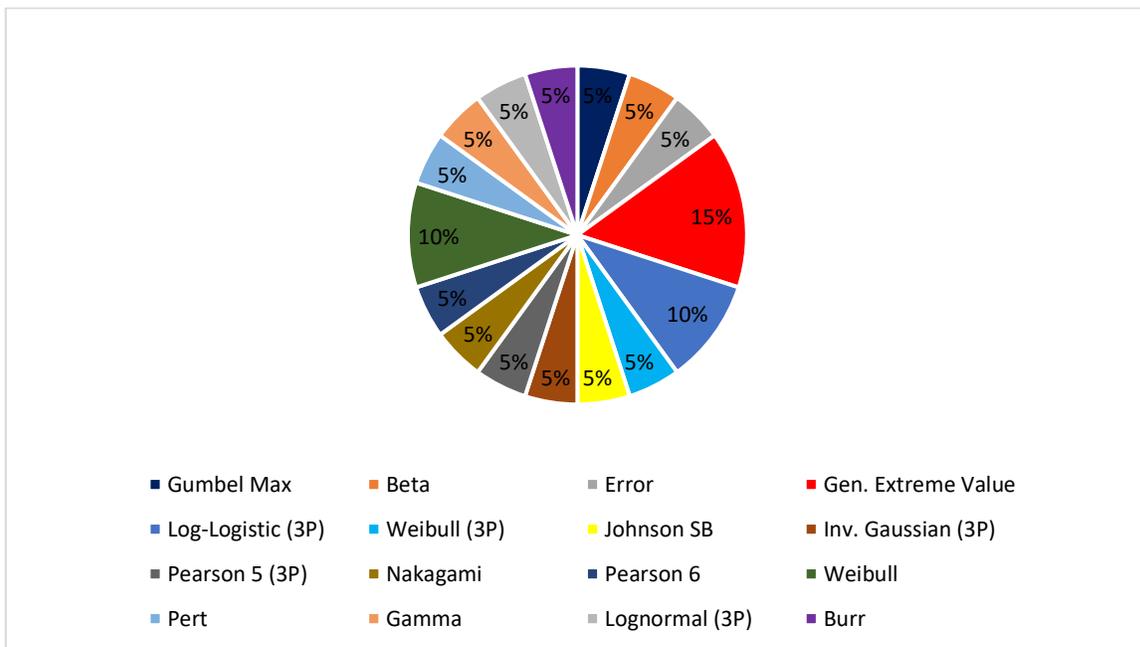


Figura 47 Distribuciones que aparecen en segunda posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación máxima en 24 horas (20).

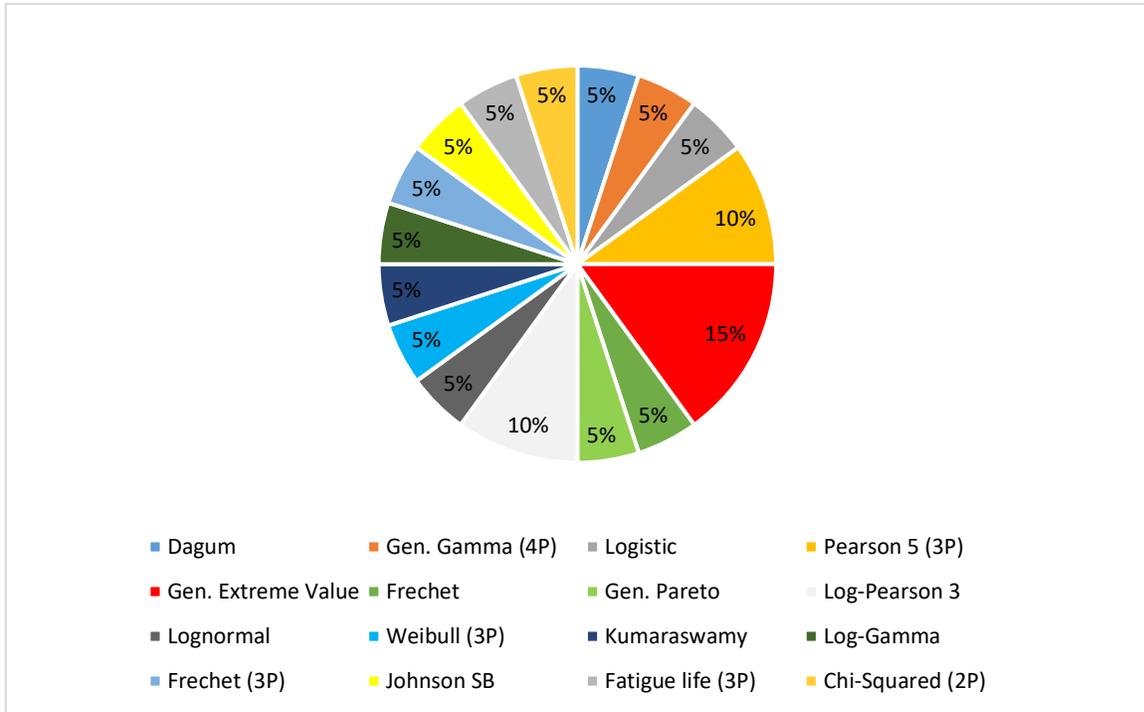


Figura 48 Distribuciones que aparecen en tercera posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación máxima en 24 horas (20).

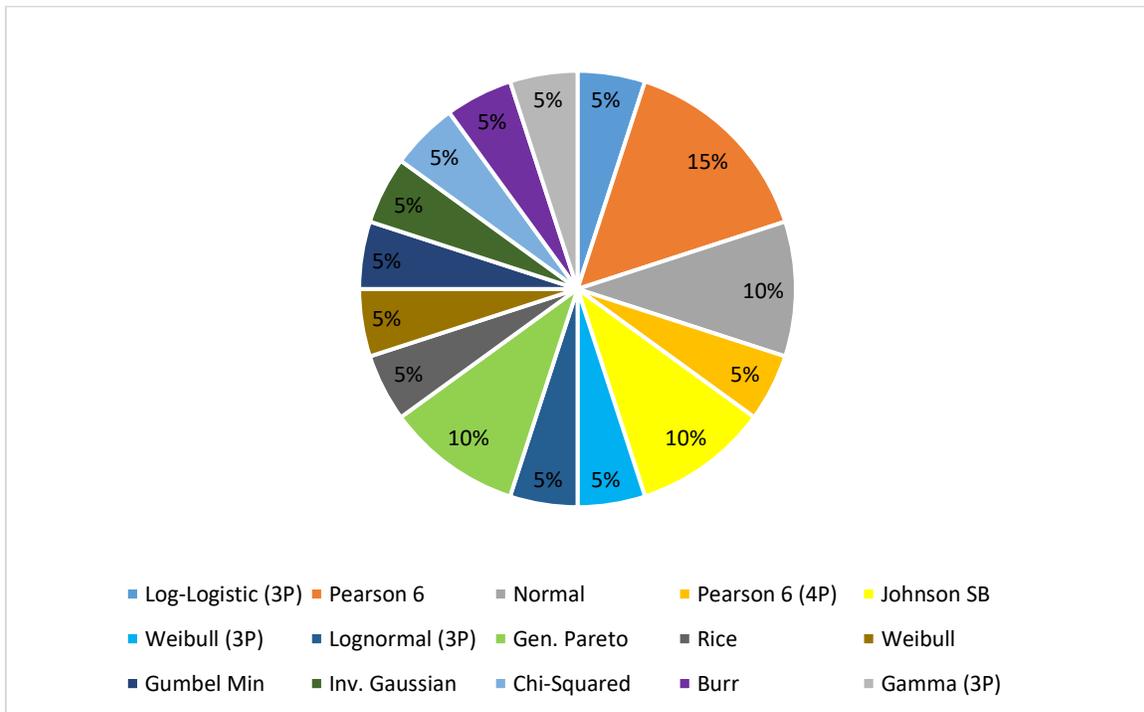


Figura 49 Distribuciones que aparecen en cuarta posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación máxima en 24 horas (20).

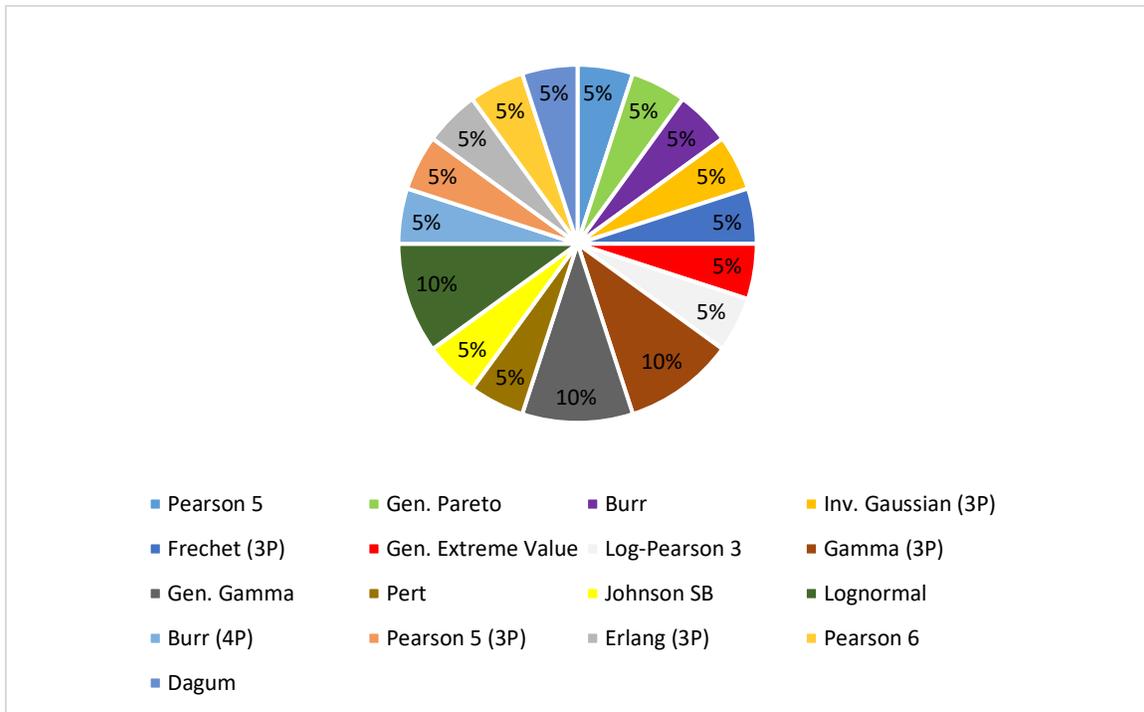


Figura 50 Distribuciones que aparecen en quinta posición del ranking según prueba de bondad de ajuste por Kolmogorov-Smirnov para precipitación máxima en 24 horas (20).

En cada figura el 100% representa un valor de 20 apariciones, por lo tanto, el 5% indica que una distribución estadística solamente apareció una vez en determinada posición. La distribución Valor extremo generalizado (identificada con color rojo) la cual apareció 7 veces en *ranking*, apareció 3 veces en la segunda posición es decir el 15% del máximo posible, en tercer lugar, apareció 3 veces con el 15%, y en quinto lugar apareció una vez con el 5%.

La distribución Johnson SB (identificada con color amarillo) la cual apareció 7 veces en *ranking*, apareció dos veces en la primera posición, mientras que en la segunda posición apareció 1 vez, en tercer lugar, apareció 1 vez con el 5%, en cuarto lugar, apareció 2 veces con el 10% y en quinto no apareció ninguna vez.

Finalmente, la distribución Log-Pearson 3 (identificada con color blanco) apareció dos veces en tercer lugar con el 10% y una vez en el quinto lugar con el 5%, mientras que Gumbel Max (identificada con color azul oscuro) obtuvo solamente una aparición en el segundo lugar con el 5%.

4.4. Precipitación máxima en 24 horas con prueba de Anderson – Darling

Se desarrolló un sistema de puntuaciones, en donde las cinco distribuciones estadísticas para cada serie de datos con mejor bondad de ajuste, recibieron un puntaje de 5 a 1, en donde 5 fue asignado a la que ocupó el primer puesto y 1 a la que ocupó el quinto puesto. La Tabla 28 presenta para cada una de las 20 estaciones seleccionadas con series anuales de precipitación máxima en 24 horas, el puntaje obtenido por la prueba de bondad de ajuste de Anderson Darling.

Tabla 28 Puntajes obtenidos por las distribuciones estadísticas (61), organizados para los 5 mejores ajustes, para cada una de las estaciones (20) seleccionadas. Series anuales de precipitación máxima en 24 horas. Prueba de Anderson-Darling

Estación	Distribución	Puntaje
ISLA DEL SANTUARIO [24015120]	Burr	5
	Dagum	4
	Log-Logistic (3P)	3
	Dagum (4P)	2
	Gen. Extreme Value	1
MERCEDES - AUT [21205660]	Gen. Gamma (4P)	5
	Johnson SB	4
	Beta	3
	Kumaraswamy	2
	Pert	1
NUNEZ [21190330]	Nakagami	5
	Burr	4
	Error	3
	Gen. Extreme Value	2
	Normal	1
CARMEN DE CARUPA [24010610]	Gen. Extreme Value	5
	Dagum	4
	Pearson 5 (3P)	3
	Lognormal (3P)	2
	Inv. Gaussian (3P)	1
SAN PABLO [23060160]	Burr	5
	Log-Logistic (3P)	4
	Frechet (3P)	3
	Gen. Extreme Value	2
	Pearson 5 (3P)	1
CUCUNUBA [24010140]	Gen. Pareto	5
	Weibull (3P)	4
	Gamma (3P)	3
	Pearson 6 (4P)	2
	Fatigue life (3P)	1
PENON EL [23060180]	Johnson SB	5
	Weibull (3P)	4
	Gamma (3P)	3
	Gen. Gamma (4P)	2
	Pearson 6 (4P)	1

Tabla 28. Continuación

Estación	Distribución	Puntaje
PINAR EL [21190310]	Gumbel Max	5
	Gen. Extreme Value	4
	Log-Pearson 3	3
	Frechet (3P)	2
	Pearson 5 (3P)	1
TUSCOLO EL [23060140]	Dagum (4P)	5
	Log-Logistic (3P)	4
	Dagum	3
	Burr	2
	Pearson 5 (3P)	1
GUASCA [21205700]	Johnson SB	5
	Pert	4
	Weibull (3P)	3
	Gamma (3P)	2
	Pearson 6 (4P)	1
SAN ISIDRO [21201610]	Johnson SB	5
	Gen. Pareto	4
	Gen. Extreme Value	3
	Gamma	2
	Log-Pearson 3	1
CABRERA LA [23065120]	Burr	5
	Pearson 5	4
	Pearson 6	3
	Log-Gamma	2
	Gen. Extreme Value	1
CHILAGUA FINCA [23060260]	Logistic	5
	Normal	4
	Error	3
	Nakagami	2
	Hypersecant	1
SAN JORGE GRANJA [21205720]	Johnson SB	5
	Weibull (3P)	4
	Burr (4P)	3
	Beta	2
	Gamma (3P)	1
SUPATA [23065210]	Log-Gamma	5
	Pearson 6	4
	Pearson 5	3
	Lognormal	2
	Fatigue life	1

Tabla 28. Continuación

Estación	Distribución	Puntaje
HATO EL [21201210]	Dagum	5
	Log-Logistic (3P)	4
	Burr	3
	Frechet (3P)	2
	Gen. Extreme Value	1
LAGUNITAS [21201190]	Erlang (3P)	5
	Gamma (3P)	4
	Weibull (3P)	3
	Fatigue life (3P)	2
	Inv. Gaussian (3P)	1
TIBACUY [21190030]	Gen. Extreme Value	5
	Gamma	4
	Gen. Gamma	3
	Log-Pearson 3	2
	Erlang	1
VILLAPINZON GRANJA [21201640]	Gamma	5
	Gen. Gamma	4
	Pearson 5 (3P)	3
	Gen. Extreme Value	2
	Log-Logistic (3P)	1
PANTANO REDONDO 1 [21201060]	Dagum	5
	Burr	4
	Log-Logistic (3P)	3
	Chi-Squared (2P)	2
	Johnson SB	1

De acuerdo con la metodología planteada en el presente trabajo de grado, el siguiente paso consistió en sumar los puntajes para cada una de las 61 distribuciones estadísticas. En la Tabla 29 se presentan los puntajes totales obtenidos para cada distribución estadística. Adicionalmente se agregó una columna con el número de veces que aparece en el *ranking* cada distribución estadística.

En color rojo se resaltan los cinco puntajes más altos y las frecuencias en el *ranking* más altas. Adicionalmente se realiza el gráfico del puntaje para cada distribución estadística en la Figura 51 y el gráfico del número de apariciones en ranking por cada distribución estadística en la Figura 52.

Tabla 29 Puntajes totales obtenidos por las distribuciones estadísticas (61). Series anuales de precipitación máxima en 24 horas. Prueba de Anderson-Darling

Distribución	Puntaje	No. Veces
Beta	5	2
Burr	28	7
Burr (4P)	3	1
Cauchy	0	0
Chi-Squared	0	0
Chi-Squared (2P)	2	1
Dagum	21	5
Dagum (4P)	7	2
Erlang	1	1
Erlang (3P)	5	1
Error	6	2
Error function	0	0
Exponential	0	0
Exponential (2P)	0	0
Fatigue life	1	1
Fatigue life (3P)	3	2
Frechet	0	0
Frechet (3P)	7	3
Gamma	11	3
Gamma (3P)	13	5
Gen. Extreme Value	26	10
Gen. Gamma	7	2
Gen. Gamma (4P)	7	2
Gen. Pareto	9	2
Gumbel Max	5	1
Gumbel Min	0	0
Hypersecant	1	1
Inv. Gaussian	0	0
Inv. Gaussian (3P)	2	2
Johnson SU	0	0
Kumaraswamy	2	1
Laplace	0	0
Levy	0	0
Levy (2P)	0	0
Log-Gamma	7	2
Log-Logistic	0	0
Log-Logistic (3P)	19	6
Log-Pearson 3	6	3
Logistic	5	1
Lognormal	2	1
Lognormal (3P)	2	1
Nakagami	7	2
Normal	5	2
Pareto	0	0

Tabla 29. Continuación

Distribución	Puntaje	No. Veces
Pareto 2	0	0
Pearson 5	7	2
Pearson 5 (3P)	9	5
Pearson 6	7	2
Pearson 6 (4P)	4	3
Pert	5	2
Power function	0	0
Rayleigh	0	0
Rayleigh (2P)	0	0
Reciprocal	0	0
Rice	0	0
Student's	0	0
Triangular	0	0
Uniform	0	0
Weibull	0	0
Weibull (3P)	18	5
Johnson SB	25	6

De la Figura 51 se observa que la distribución Burr es la que mejor puntaje obtuvo con 28 puntos, la segunda fue Valor extremo generalizado con un puntaje de 26, la tercera fue Johnson SB con un puntaje de 25, la cuarta Dagum con 21 puntos y la quinta fue Log-logistic (3P) con 19 puntos. También se observa que 37 distribuciones estadísticas tuvieron un puntaje superior a 1, lo cual representa el 60.6% de las distribuciones estudiadas, mientras que 24 distribuciones que representan el 39.4% de las distribuciones estudiadas no obtuvo ningún puntaje. Por otra parte, la distribución Valor extremo generalizado y Gumbel, las cuales son referente para valores extremos obtuvieron un puntaje de 26 y 5 respectivamente, y se ubicaron en los puestos 2 y 23 respectivamente. Log-Pearson 3 la cual también es referente a nivel mundial para valores extremos, obtuvo un puntaje de 6 y un puesto de 20.

De la Figura 52 se observa que la distribución Valor extremo generalizada es la que más veces aparece en el *ranking* con 10 apariciones, la segunda fue Burr con 7 apariciones, la tercera y cuarta fueron Log-logistic (3P) y Johnson SB con 6 apariciones, y la quinta fue Dagum con 5 apariciones. También se observa que 37 distribuciones estadísticas tuvieron al menos una aparición en el *ranking*, lo cual representa el 60.6% de las distribuciones estadísticas estudiadas, mientras que 24 distribuciones que representan el 39.4% de las distribuciones estadísticas estudiadas no tuvieron ninguna aparición en *ranking*. Por otra parte, las distribuciones Valor extremo generalizado, Gumbel y Log-Pearson 3 las cuales son referente para valores extremos, obtuvieron 10,1 y 3 apariciones en el *ranking* ubicándose en los puestos 1, 32 y 11 respectivamente.

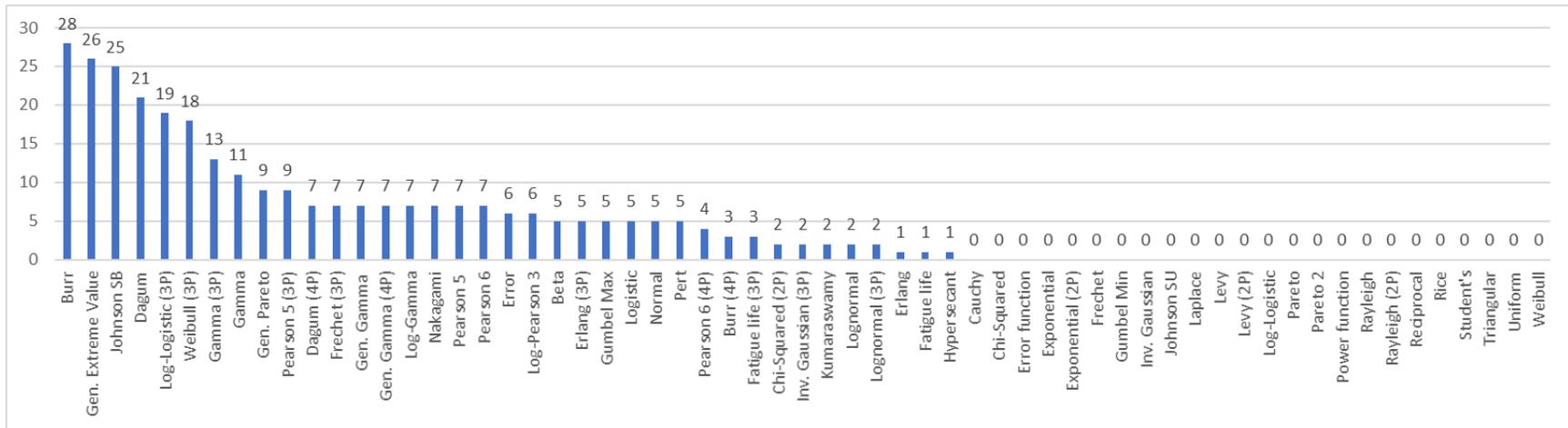


Figura 51 Puntajes totales obtenidos por las distribuciones estadísticas (61) para las series de precipitación máxima en 24 horas. Prueba de Anderson-Darling

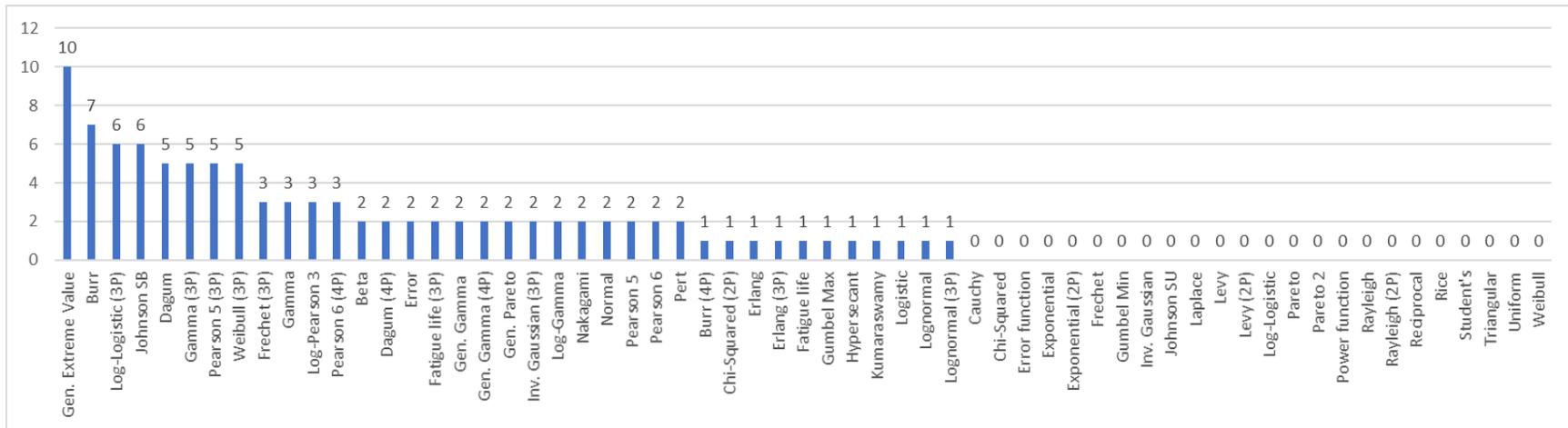


Figura 52 Número de apariciones en el ranking por las distribuciones estadísticas (61) para las series de precipitación máxima en 24 horas. Prueba de Anderson-Darling

En la Tabla 30 y en la Figura 53 se presentan los puntajes obtenidos para las cinco mejores distribuciones estadísticas según la prueba de bondad de ajuste por Anderson-Darling. La línea roja punteada representa el puntaje máximo posible, el cual tiene un valor de 100 y representa el puntaje máximo que obtendría la distribución estadística ideal si llegara a ocupar el primer puesto en las 20 series de datos analizadas.

Tabla 30 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Anderson-Darling

Distribución estadística	Puntaje
Burr	28
Gen. Extreme Value	26
Johnson SB	25
Dagum	21
Log-Logistic (3P)	19
Log-Pearson 3	6
Gumbel Max	5

Notas: Los puntajes para cada serie de datos se tomaron de 5 a 1. En este cuadro se han incluido las 5 distribuciones con mejor puntaje y las distribuciones Log-Pearson 3 y Gumbel de referencia. La distribución Burr apareció 3 veces en el primer lugar, y la distribución GEV 2 veces en el primer lugar.

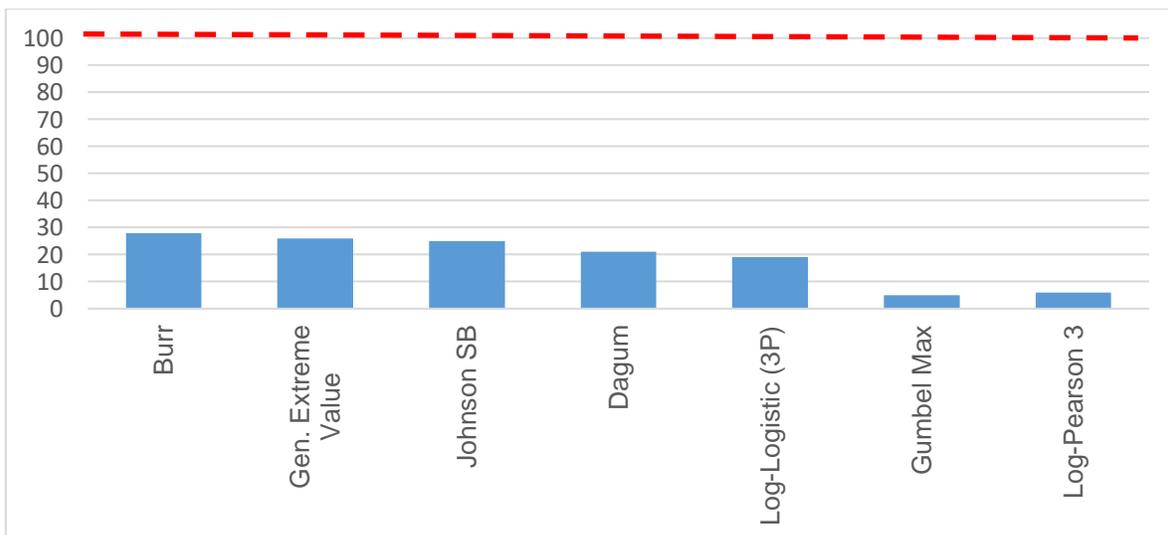


Figura 53 Puntajes obtenidos por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Anderson-Darling. Distribuciones Gumbel y Log Pearson 3 de referencia

En el primer lugar Burr obtuvo un puntaje de 28 sobre 100 máximos posibles, en segundo lugar, valor extremo generalizado obtuvo un puntaje de 26 el cual no tiene gran diferencia con respecto al primero. Por otra parte, de las distribuciones que son referente para ajustar valores extremos, Log-Pearson 3 obtuvo 6 puntos, es decir 4.6 veces menos puntaje en comparación con los primeros puestos y Gumbel obtuvo 5 puntos, es decir 5.6 veces menos puntaje en comparación con los primeros puestos.

En la Tabla 31 y en la Figura 54 se presenta el número de apariciones para las cinco mejores distribuciones estadísticas según la prueba de bondad de ajuste por Anderson-Darling. La línea roja punteada representa la cantidad máxima de apariciones posible, la cual tiene un valor de 20 y representa la cantidad máxima de apariciones que tendría la distribución estadística ideal si llegara a aparecer dentro de las cinco mejores distribuciones estadísticas para las 20 series de datos analizadas.

Tabla 31 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Anderson-Darling

Distribución estadística	No. Veces en ranking
Gen. Extreme Value	10
Burr	7
Log-Logistic (3P)	6
Johnson SB	6
Dagum	5
Log-Pearson 3	3
Gumbel Max	1

Notas: Los puntajes para cada serie de datos se tomaron de 5 a 1.

En este cuadro se han incluido las 5 distribuciones con mejor puntaje y las distribuciones Log-Pearson 3 y Gumbel de referencia.

La distribución Burr apareció 3 veces en el primer lugar, y la distribución GEV 2 veces en el primer lugar.

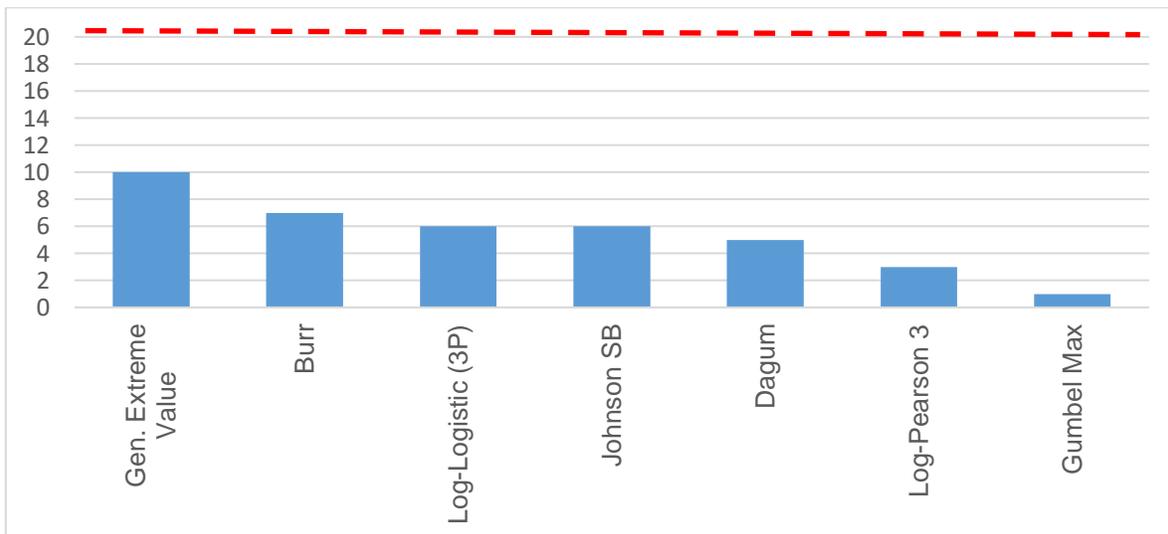


Figura 54 Número de apariciones en el ranking por diferentes distribuciones estadísticas (61), para series de precipitaciones máximas en 24 horas (20). Prueba de Anderson-Darling. Distribuciones Gumbel y Log Pearson 3 de referencia

Valor extremo generalizado obtuvo 10 apariciones de 20 máximas posibles, y Johnson SB obtuvo 7 apariciones de 20 máximas posibles, este resultado coincide con los puntajes más altos y las ubica en primera y segunda posición. Por otra parte, la distribución Log-Pearson 3 obtuvo 3 apariciones, lo que equivale a 3.3 veces menos con respecto a la que ocupó el primer puesto y Gumbel Max obtuvo una aparición, es decir 10 veces menos en comparación con la que ocupó el primer lugar.

Desde la Figura 55 hasta la Figura 59 se presentan los porcentajes de apariciones en el *ranking* de las distribuciones estadísticas que ocuparon los primeros, segundos, terceros, cuartos y quintos lugares respectivamente.

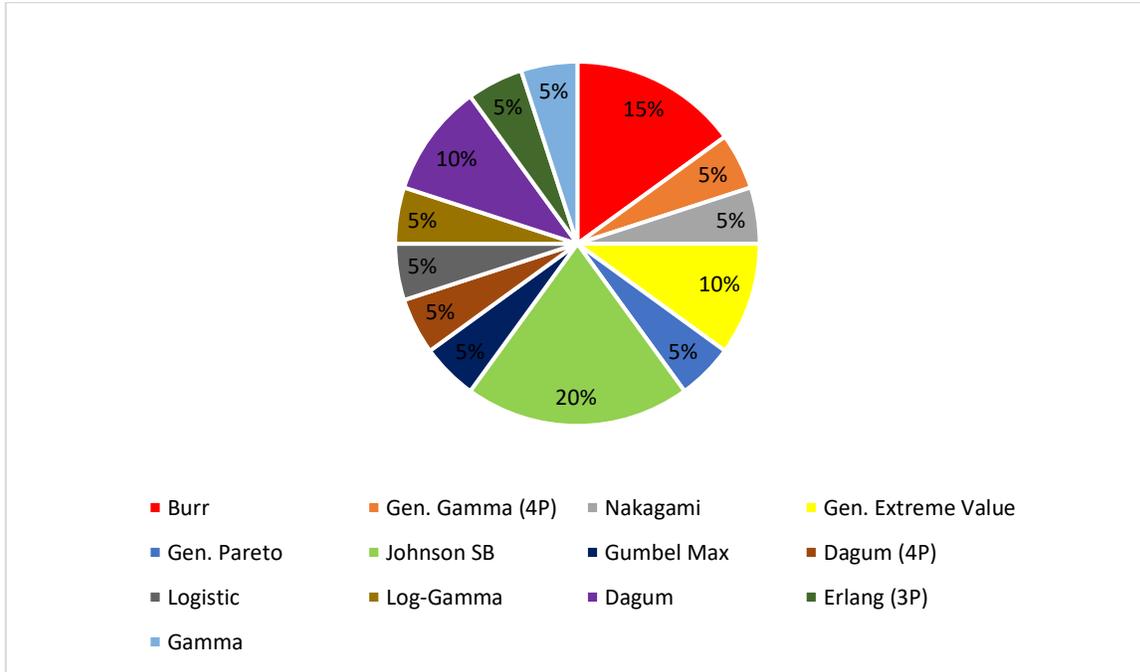


Figura 55 Distribuciones que aparecen en primera posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación máxima en 24 horas (20).

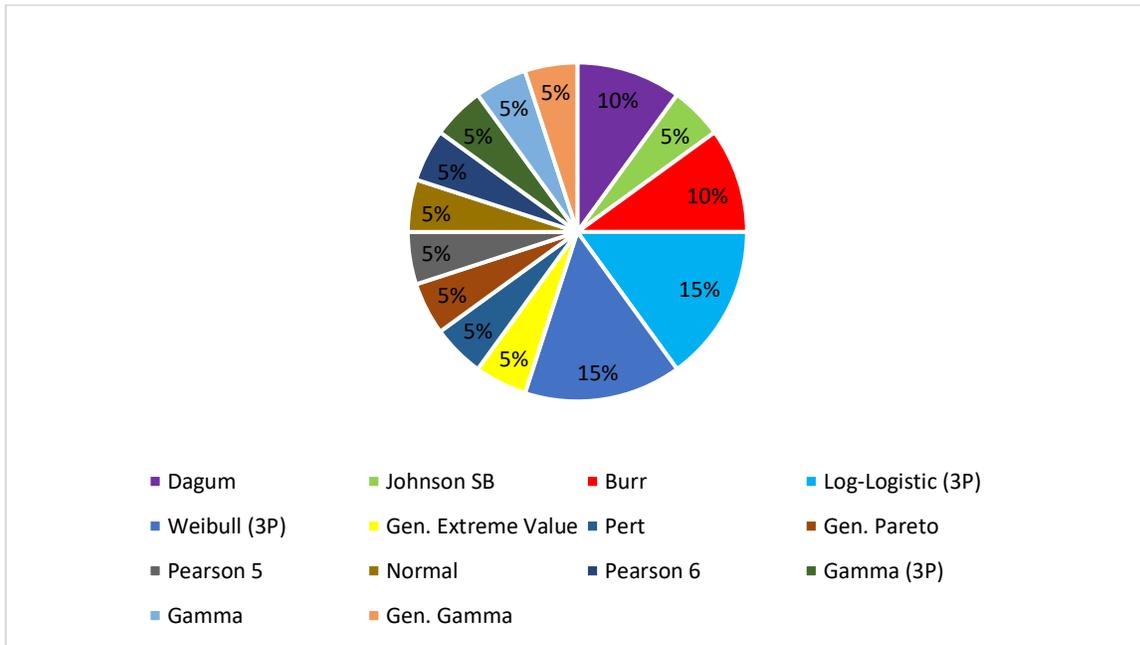


Figura 56 Distribuciones que aparecen en segunda posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación máxima en 24 horas (20).

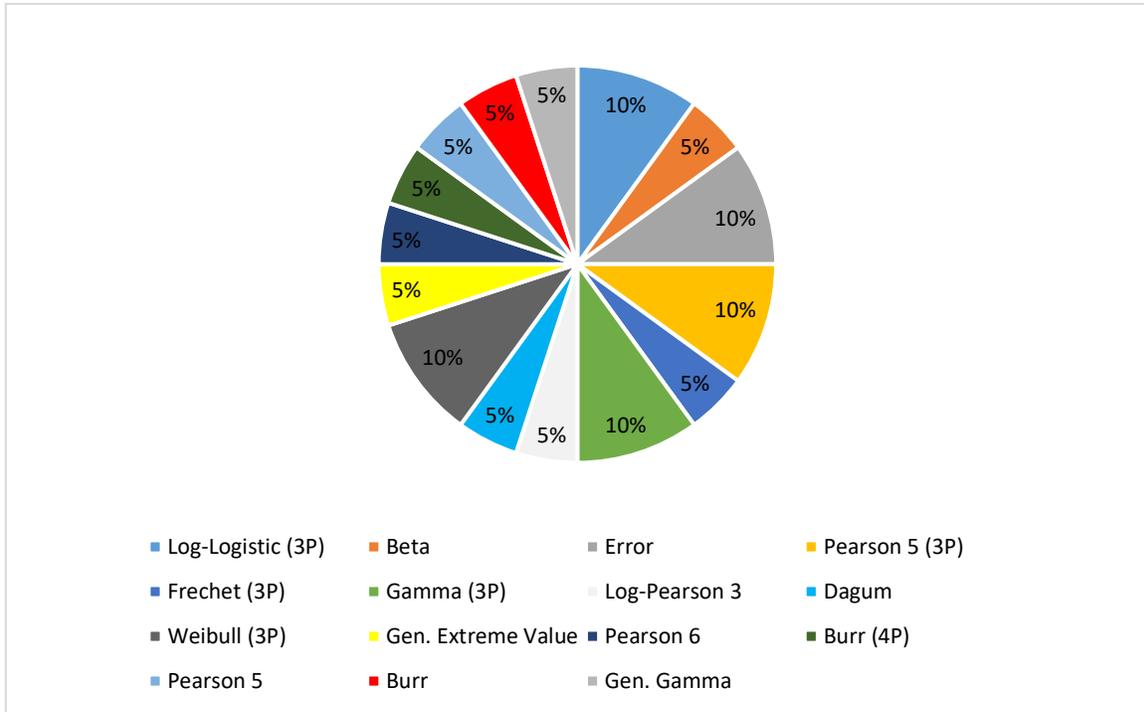


Figura 57 Distribuciones que aparecen en tercera posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación máxima en 24 horas (20).

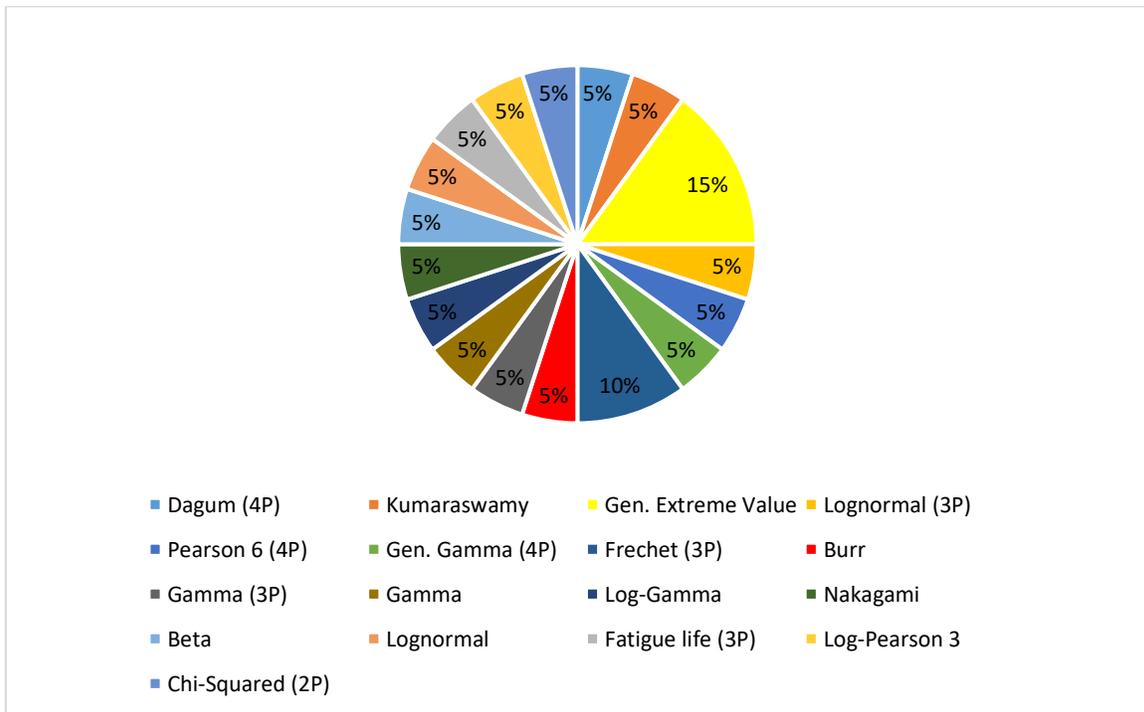


Figura 58 Distribuciones que aparecen en cuarta posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación máxima en 24 horas (20).

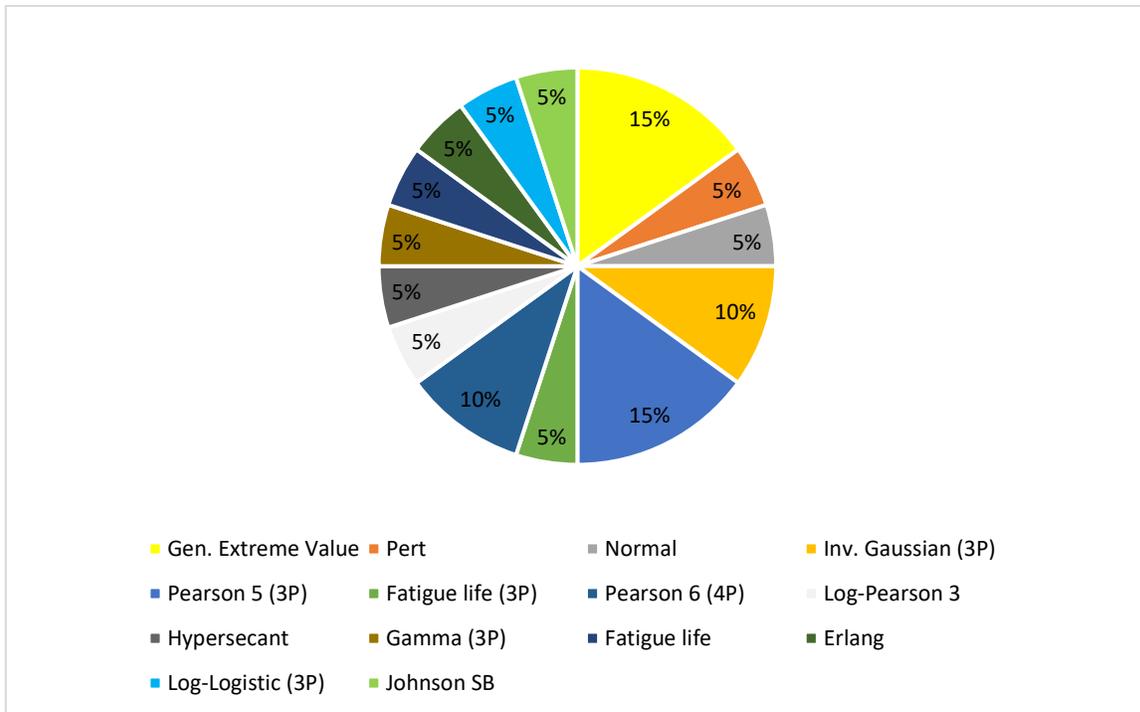


Figura 59 Distribuciones que aparecen en quinta posición del ranking según prueba de bondad de ajuste por Anderson-Darling para precipitación máxima en 24 horas (20).

En cada figura el 100% representa un valor de 20 apariciones, por lo tanto, el 5% indica que una distribución estadística solamente apareció una vez en determinada posición. La distribución Valor extremo generalizado (identificada con color amarillo) la cual apareció 10 veces en *ranking*, apareció 2 veces en primer lugar con el 10%, 1 vez en la segunda posición con el 5% del máximo posible, en tercer lugar, apareció 1 vez con el 5%, en cuarto lugar, apareció 3 veces con el 15% y en quinto lugar apareció tres veces con el 15%.

La distribución Burr (identificada con color rojo) la cual apareció 7 veces en *ranking*, apareció tres veces en la primera posición con el 15%, mientras que en la segunda posición apareció 2 veces con el 10%, en tercer lugar, apareció 1 vez con el 5%, en cuarto lugar, apareció 1 vez con el 5% y en quinto no apareció ninguna vez.

Finalmente, la distribución Log-Pearson 3 (identificada con color blanco) apareció una vez en tercer lugar con el 5%, una vez en cuarto lugar con el 5% y una vez en el quinto lugar con el 5%, mientras que Gumbel Max (identificada con color azul oscuro) obtuvo solamente una aparición en el primer lugar con el 5%.

5. CONCLUSIONES Y RECOMENDACIONES

- Las distribuciones estadísticas más utilizadas en hidrología para ajustar valores máximos son Log-Pearson 3 (Norteamérica) y Gumbel (Europa), y para valores medios es la distribución normal. Teniendo en cuenta que la dinámica de los procesos que intervienen en el ciclo hidrológico es única en cada región geográfica del planeta, se confirmó si las distribuciones estadísticas más utilizadas en hidrología eran adecuadas al aplicarlas en series de datos en una zona específica de Cundinamarca.
- Se hizo la revisión bibliográfica de diferentes estudios donde se aplicaron varias distribuciones estadísticas para ajustar series de datos de valores medios y extremos de precipitación en diferentes países, y se encontró que hay distribuciones estadísticas que no son de uso común en hidrología pero que tienen mejores bondades de ajuste en comparación con las convencionales.
- Se seleccionaron 20 series de datos de precipitación total anual y 20 series de datos de precipitación máxima en 24 horas con registros superiores a 30 años y que cumplieron con propiedades tales como homogeneidad, estacionariedad, no tendencia y no autocorrelación. La no autocorrelación permite verificar que los datos provienen de un proceso aleatorio, la homogeneidad y la estacionariedad de las series de datos permiten establecer que la muestra proviene de una misma población, también que la media y la varianza son estables en el tiempo y adicionalmente con la falta de tendencia se pueden aplicar de manera adecuada los modelos probabilísticos, ya que su filosofía está basada en que los eventos que sucedieron en el pasado sucederán de la misma forma en el futuro.
- Se realizó el ajuste de cada serie de datos a 61 distribuciones estadísticas, y se realizó la prueba de bondad de ajuste a cada una, aplicando las pruebas de Kolmogorov-Smirnov y Anderson-Darling. Estas pruebas fueron seleccionadas porque de acuerdo con las referencias consultadas, diferentes estudios han demostrado que son las que presentan mayor sensibilidad en series de datos de 30 a 50 valores.
- Se desarrolló un sistema de puntuaciones, en donde las cinco distribuciones estadísticas para cada serie de datos con mejor bondad de ajuste, recibieron un puntaje de 5 a 1, en donde 5 fue asignado a la que ocupó el primer puesto y 1 a la que ocupó el quinto puesto. Por último, se sumaron los puntajes totales obtenidos por cada distribución estadística y se identificaron las que mejor puntaje obtuvieron siendo 100 el puntaje máximo posible (Ver tablas 18,22,26 y 30). El puntaje máximo posible representa el puntaje que obtendría la

distribución estadística ideal si llegara a ocupar el primer puesto en las 20 series de datos analizadas. Del mismo modo, la cantidad máxima que pudo aparecer una distribución estadística en el Ranking fue de 20 veces.

- La metodología propuesta en el trabajo de grado, se creó y desarrolló para una zona particular del departamento de Cundinamarca. No obstante, se puede generalizar y aplicar en otras regiones o en todo el país, y se puede extender a otras variables hidrológicas.
- De acuerdo con los resultados, para la variable de precipitación total anual, al realizar las pruebas de bondad de ajuste por Kolmogorov-Smirnov con 61 distribuciones estadísticas, 27 no obtuvieron ningún puntaje, lo cual representa el 44.3% de las distribuciones analizadas. Para la prueba de Anderson-Darling, 32 distribuciones no obtuvieron ningún puntaje, lo cual representa el 52.4% de las distribuciones estadísticas analizadas.
- Para la variable de precipitación total anual, según la prueba de bondad de ajuste por Kolmogorov-Smirnov, la distribución Log-Logistic (3P) obtuvo 44 puntos de 100 máximos posibles y ocupó el primer lugar, mientras que Burr ocupó el segundo lugar con 21 puntos, es decir casi la mitad del puntaje con respecto a Log-Logistic (3P). Por otra parte, la distribución normal obtuvo un puntaje de 8, se podría afirmar que Log-logistic (3P) obtuvo un desempeño 5.5 veces mayor en comparación con la distribución normal y es la distribución estadística que mejor bondad de ajuste tiene en la región estudiada de acuerdo con el puntaje establecido.
- Para la variable de precipitación total anual, según la prueba de bondad de ajuste por Anderson-Darling, la distribución Log-Logistic (3P) obtuvo 43 puntos de 100 máximos posibles y ocupó el primer lugar, mientras que Burr ocupó el segundo lugar con 25 puntos, es decir más de la mitad del puntaje con respecto a Log-Logistic (3P). Por otra parte, la distribución normal obtuvo un puntaje de 5, se podría afirmar que Log-logistic (3P) obtuvo un desempeño 8.6 veces mayor en comparación con la distribución normal y al igual que en la prueba de Kolmogorov-Smirnov, es la distribución estadística que mejor bondad de ajuste tuvo en la región estudiada de acuerdo con el puntaje establecido.
- De acuerdo con los resultados, para la variable de precipitación máxima en 24 horas, al realizar las pruebas de bondad de ajuste por Kolmogorov-Smirnov con 61 distribuciones estadísticas, 20 no obtuvieron ningún puntaje, lo cual representa el 32.7% de las distribuciones analizadas. Para la prueba de Anderson-Darling, 24 distribuciones no obtuvieron ningún puntaje, lo cual representa el 39.3% de las distribuciones estadísticas analizadas.

- Para la variable de precipitación máxima en 24 horas, según la prueba de bondad de ajuste por Kolmogorov-Smirnov, las distribuciones valor extremo generalizado y Johnson SB obtuvieron 22 puntos de 100 máximos posibles cada una, lo cual las ubica en el primer lugar. El tercer lugar fue para la distribución Pareto Generalizada la cual obtuvo 18 puntos, es decir 0.81 veces el puntaje en comparación con el primer puesto. Por otra parte, de las distribuciones que son referente para ajustar valores extremos, Log-Pearson 3 obtuvo 7 puntos, es decir casi 3 veces menos puntaje en comparación con los primeros puestos y Gumbel obtuvo 4 puntos, es decir 5.5 veces menos puntaje en comparación con los primeros puestos. Valor extremo generalizado y Johnson SB son las distribuciones que mejor ajuste presentaron de acuerdo con el puntaje establecido.
- Para la variable de precipitación máxima en 24 horas, según la prueba de bondad de ajuste por Anderson-Darling, la distribución de Burr obtuvo un puntaje de 28 sobre 100 máximos posibles ocupando el primer lugar, en segundo lugar, valor extremo generalizado obtuvo un puntaje de 26 el cual no tiene gran diferencia con respecto al primero. El tercer puesto fue para la distribución Johnson SB la cual obtuvo 25 puntos. Por otra parte, de las distribuciones que son referente para ajustar valores extremos, Log-Pearson 3 obtuvo 6 puntos, es decir 4.6 veces menos puntaje en comparación con los primeros puestos y Gumbel obtuvo 5 puntos, es decir 5.6 veces menos puntaje en comparación con los primeros puestos. Burr, valor extremo generalizado y Johnson SB son las distribuciones que mejor ajuste presentaron de acuerdo con el puntaje establecido.
- Para estudios posteriores, se recomienda comparar gráficamente para cada serie de datos de cada variable hidrológica la distribución estadística de referencia contra la distribución estadística que ocupe el primer puesto. Lo anterior con el fin de detectar de manera visual las diferencias con respecto a la mejor bondad de ajuste presentada.

6. REFERENCIAS BIBLIOGRÁFICAS

- Abida, H., & Ellouze, M. (2008). Probability distribution of flood flows in Tunisia. *Hydrology and earth system sciences*, 12(1), 703-714.
- Adeloye, A. J., & Montaseri, M. (2002). Preliminary streamflow data analyses prior to water resources planning study. *Hydrological Sciences–Journal–des Sciences Hydrologiques*, 679-692.
- Aguilera, M. (2007). *Estimación de funciones de distribución de probabilidad, para caudales máximos, en la region de Maule*. Talca, Chile: Universidad de Talca.
- Ahmad, M. I., Sinclair, C. D., & Spurr, B. D. (1988). Assessment of flood frequency models using empirical distribution function statistics. *Water Resources Research*, 1323-1328.
- Ahsanullah, M., & Nevzorov, V. B. (2014). Some Inferences on the Levy distribution. *Journal of Statistical Theory and Applications*, 205-211.
- Alam, M., Emura, K., Farnham, C., & Yuan, J. (2018). Best-Fit probability distributions and period returns for maximum monthly rainfall in Bangladesh. *Climate*, 6(9), 1-16, doi: 10.3390/cli6010009.
- Alam, S., Khan, M., & Rahat, S. (2015). A study on selection of probability distributions of extreme hydrologic parameters for the peripheral river system in Dhaka City. *Proceedings of Academics World 15th International Conference*. Bangkok: Bangladesh University of Technology.
- Alfaro, E. J., & Soley, J. F. (2009). Descripción de dos métodos de rellenado de datos ausentes en series de tiempo meteorológicas. *Revista de matemática: Teoría y Aplicaciones*, 60-75.
- Amin, M., Rizwan, M., & Alazba, A. (2015). A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan. *Open life Sc*, 2016(11), 432-440, doi: 10.1515/boil-2016-0057.
- Benson, M. A. (1968). Uniform flood-frequency estimating methods for federal agencies. *Water Resources Research*, 891-908.
- Bethea, R. M., & Rhinehart, R. R. (1994). Applied Engineering Statistics. *Technometrics*, 69-95.
- Bose, P., & Morin, P. (2002). Algorithms and Computation. *13th International Symposium*. Berlin: Springer.

- Box, G. E., & Jenkins, G. M. (1976). *Time series analysis, forecasting and control*. San Francisco: Holden Day.
- Brissette, F., & Chen, J. (2013). Finding the most appropriate precipitation probability distribution for stochastic weather generation and hydrological modelling in Nordic watersheds. *Hydrological Processes*, 3718-3729.
- Burke, E. J., Perry, H. J., & Brown, S. J. (2010). An extreme value analysis of UK drought and projections of change in the future. *Journal of Hydrology*, 131-151.
- Cao, L., Zhong, J., Su, B., Zhai, J., & Gemmer, M. (2013). Probability distribution and projected trends of daily precipitation in China. *Advances in climate change research*, 4(3), 153-159.
- Chow V. T. (1993). *Hidrología aplicada*. Bogotá: McGraw-Hill.
- Dahmen, E. R., & Hall, M. J. (1990). *Screening of hydrological data*. Wageningen: ILRI.
- Durrett, R. (2010). *Probability*. Cambridge: Cambridge University Press.
- Esberto, D. (2018). Probability distribution fitting of rainfall patterns in Philippine regions for effective risk management. *Environment and ecology research*, 6(3), 178-186, doi: 10.13189/eer.2018.060305.
- Fischer, T., Su, B., Luo, Y., & Scholten, T. (2012). Probability distribution of precipitation extremes for weather index-based insurance in the Zhujiang river basin, South China. *Journal of hydrometeorology*, 13(1), 1023-1036, doi: 10.1175/JHM-D-11-041.1.
- Fisher, R. A., & Tippett, L. H. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 180-192.
- Freeze, A., & Cherry, T. A. (1979). *Groundwater*. New Jersey: Prentice Hall.
- Gado, T., & Zeidan, B. (2018). On selection of probability distributions for annual extreme rainfall series in Egypt. *Twenty-first international water technology conference IWTC-21, Ismailia*. Tanta University Egypt.
- García, A. (2013). *Análisis de distribuciones estadísticas alternativas a las tradicionales para la optimización de los caudales de cálculo empleados en los estudios hidrológicos*. Universidad de Extremadura, Badajoz, España.: Tesis Doctoral.
- Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. New York: John Wiley and Sons.

- Hanson, L., & Vogel, R. (2008). The probability distribution of daily rainfall in the United States. *World environmental and water resources congress*, (págs. 1-12). Honolulu- Hawaii.
- Hosking, J. R., & Wallis, J. R. (1997). *Regional Frequency Analysis*. Cambridge: Cambridge University Press.
- Hosking, R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society*, 105-124.
- Izinyon, C., & Iyekoetin, E. (2017). Best-Fit probability distribution model for rainfall frequency analysis of three cities in south eastern Nigeria. *Nigerian journal of environmental sciences and technology (NIJETS)*, 1(1), 34-42.
- Kavvas, M. L., & Delleur, J. W. (1981). A stochastic cluster model of daily rainfall sequences. *Water Resources Research*, 1151-1160.
- Kim, D., Lee, S., Hong, Y., & Lee, E. (2010). The determination of probability distributions of Annual, seasonal and monthly precipitation in Korea. *Korean journal of agricultural and forest meteorology*, 12(2), 83-94.
- Koschinchanov, G., & Dimitrov, D. (2009). Precipitation intensity probability distribution modelling for hydrological and construction design purposes. *IOP conference series: Earth and environmental sciences*, 4(1), 1-12, doi: 10.1088/1755-1307/4/4/012009.
- Kumar, D. (2017). The Burr type XII distribution with some statistical properties. *Journal of Data Science*, 509-534.
- Law, A. M., & Kelton, W. D. (1997). *Simulation Modeling & Analysis*. New York: McGraw-Hill.
- Lehoczky, J., & Schervish, M. (2018). Overview and History of Statistics for Equity Markets. *Annual Review of Statistics and Its Application*, 265-288.
- Li, Z., Li, Z., Zhao, W., & Wang, Y. (2015). Probability modeling of precipitation extremes over two river basins in northwest of China. *Advances in Meteorology*, 2015(1), 1-13.
- Machiwal, D., & Kumar, M. J. (2012). *Hydrologic time series analysis*. Dordrecht: Springer.
- Mahgoub, T., & Ibrahim, A. (2015). Fitting probability distributions of annual rainfall in Sudan. *SUST Journal of engineering and computer sciences (JECS)*, 17(2), 2-7.
- Maidment, D. R. (1997). *Handbook of hydrology*. New York: McGraw-Hill.
- Mamoon, A., & Rahman, A. (2017). Selection of the best fit probability distribution in rainfall frequency analysis for Qatar. *Natural Hazards*, 86(1), 281-296.

- Markovic, R. (1965). Probability functions of best fit to distributions of annual precipitation and runoff. (C. s. university, Ed.) *Hydrology Papers*, 8(1), 1-33.
- Monjo, R., Caselles, V., & Chust, G. (2012). Alternative model for precipitation probability distribution: Application to Spain. *Climate research*, 1(1), 23-33, doi: 10.3354/cr01055.
- Monsalve G. (1995). *Hidrología en la ingeniería*. Bogotá: Alfaomega.
- MOP. (2012). manual de carreteras. Santiago de Chile .
- Mora, A. (2011). Estudio comparativo de algunos estimadores del índice de cola . *E. INNOVAR.* , 17-33.
- Oosterbaan, R. J. (1994). *Drainage Principles and Applications, Capítulo 6*. Wageningen: International Institute for Land Reclamation and Improvement.
- Pilon, P. J., & Harvey , K. D. (1994). Consolidated frequency analysis. *Environment Canada*, 137-147.
- Pizarro, R., González , P., Witterssheim, M., Saavedra, J., & Soto, C. (1993). *Elementos técnicos de hidrología III*. Talca: Universidad de Talca.
- Planos E. O. (2015). Influencia de las grandes precipitaciones en la garantía de entrega de agua anual. *Aqua-LAC*, 38-48.
- Ritzema , H. P. (1979). *Drainage principles and applications*. Wageningen: International Institute for Land Reclamation and Improvement.
- ROSTLAC. (1982). *Guía metodológica para la elaboración del balance hídrico de América del sur*. Montevideo: UNESCO.
- Shoukri, M. M. (1988). Sampling properties of estimators of the log-logistic distribution with application to Canadian precipitation data. *The Canadian Journal of Statistics*, 223-236.
- Stephens, M. A. (1979). Tests of Fit for the Logistic Distribution Based on the Empirical Distribution Function. *Biometrika*, 591-595.
- Stephenson D.J. (1981). *Stormwater hydrology and Dranaige*. Amsterdam: Elsevier Scientific Pub.
- Sukrutha, A., & Desai, S. (2017). Probability distribution for monthly precipitation data in India. *Research archive for indian institute of technology Hyderabad RAIITH*, 1(1), 1-23.
- Thanh, N. (2017). Fitting probability distribution to extreme precipitation for a limited mountain area in Vietnam. *Journal of geoscience and environment protection*, 5(1), 92-107.

- Vieira, F., Casarim, J., Vismara, E., & Possenti, J. (2018). Probability distributions of frequency analysis of rainfall at the southwest region of Paraná state, Brazil. *Revista de ciências agroveterinárias*, 17(2), 260-266.
- Weibull, W. (1939). A Statistical Theory of the Strength of Materials. *Ingeniörsvetenskapsakademien*, 50-95.
- Yue, S., & Hashino, M. (2007). Probability distribution of annual, seasonal and monthly precipitation in Japan. *Hydrological sciences journal*, 52(5), 863-877, doi: 10.1623/hysj.52.5.863.
- Zaman, M., Rahman, A., Haddad, K., & Hagare, D. (2013). A study on selection of probability distributions for at-site flood frequency analysis in Australia. *Natural Hazards*, 69(3), 1803-1813.
- Zeng, X., Wang, D., & Wu, J. (2015). Evaluating the Three Methods of Goodness of Fit Test for Frequency Analysis. *Journal of Risk Analysis and Crisis Response*, 178-187.