

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

**ESCUELA DE POSGRADO
MAESTRÍA EN RECURSOS HÍDRICOS**



**“ANÁLISIS DE SEQUÍAS USANDO SERIES SINTÉTICAS
EN LA CUENCA DEL RÍO MANTARO”**

Presentada por:

JOSÉ ANTONIO CAIRAMPOMA AMARO

**TESIS PARA OPTAR EL GRADO DE
MAGISTER SCIENTIAE RECURSOS HÍDRICOS**

Lima - Perú

2021

UNIVERSIDAD NACIONAL AGRARIA

LA MOLINA

ESCUELA DE POSGRADO

MAESTRÍA EN RECURSOS HÍDRICOS

**“ANÁLISIS DE SEQUÍAS USANDO SERIES SINTÉTICAS EN LA
CUENCA DEL RÍO MANTARO”**

**TESIS PARA OPTAR EL GRADO DE
MAGISTER SCIENTIAE**

Presentada por:

JOSÉ ANTONIO CAIRAMPOMA AMARO

Sustentada y aprobada ante el siguiente jurado:

.....
Ph.D. Eduardo Chávarri Velarde
PRESIDENTE

.....
Mg.Sc. Gonzalo Fano Miranda
ASESOR

.....
Dr. Abel Mejía Marcacuzco
MIEMBRO

.....
Mg.Sc. Ricardo Apaella Nalvarte
MIEMBRO

DEDICATORIA

A mi familia y a todos quienes me dieron el aliento para no declinar en el intento y poder terminar la presente tesis, en especial a mis dos hijas, Margaret y Jolie, que son la razón de mi ser.

AGRADECIMIENTOS

A la Facultad de Agronomía de la Universidad Nacional del Centro del Perú, por haberme dado la oportunidad de seguir con los estudios de maestría y a los docentes de la maestría de Recursos Hídricos de la Universidad Nacional Agraria la Molina, por haberme impartido los conocimientos para poder alcanzar mi desarrollo profesional y también personal.

ÍNDICE GENERAL

I.	INTRODUCCIÓN	1
1.1	OBJETIVOS DE LA INVESTIGACIÓN	2
1.1.1	Objetivos Generales	2
1.1.2	Objetivos específicos	3
II.	REVISIÓN DE LITERATURA	4
2.1	SEQUÍAS	4
2.2	DEFINICIÓN DE SEQUÍAS	4
2.3	TIPOS DE SEQUÍAS	5
2.3.1	Sequía meteorológica	5
2.3.2	Sequía hidrológica	6
2.3.3	Sequía agrícola	6
2.3.4	Sequía socioeconómica	6
2.4	ANÁLISIS ESTOCÁSTICO DE SEQUÍAS	7
2.4.1	Análisis e identificación de sequías	7
2.4.2	Caracterización de sequías	7
2.5	ÍNDICES DE SEQUÍAS	9
2.5.1	Índice de Precipitación Estandarizado	9
2.5.2	Cálculo de ISP	9
2.6	CARACTERIZACIÓN DE LAS SEQUÍAS	12
2.7	SERIES TEMPORALES Y SU ANÁLISIS	12
2.7.1	Test de homogeneidad o consistencia de las series de tiempo	13
2.7.2	Ratio de Von Neumann	14
2.7.3	Test de Buishand	14
2.7.4	Prueba estándar de homogeneidad normal	15
2.7.5	Prueba de Pettitt	16
2.7.6	Análisis de la estacionariedad	17
2.7.7	Análisis de tendencias Mann - Kendall	17
2.8	GENERACIÓN DE SERIES SINTÉTICAS	20
2.8.1	Procesos autorregresivos (AR)	20
2.8.2	Procesos de medias móviles	21
2.8.3	Proceso autorregresivo de medias móviles (ARMA)	22
2.8.4	Metodología para la aplicación del modelo ARMA	22
2.9	ANÁLISIS UNIVARIADO DE SEQUÍAS	26
2.10	ESTIMACIÓN DE PARÁMETROS	27

2.11	MODELOS UNIVARIADAS.....	27
2.11.1	Análisis multivariado de sequías.....	31
2.11.2	Cóputas.....	31
2.11.3	Cóputas arquimedianas	31
2.12	PERIODO DE RETORNO Y RIESGO	32
2.12.1	Evento	33
2.12.2	Intervalo de ocurrencia.....	33
2.12.3	Periodo de retorno	33
2.13	PERIODO DE RETORNO DE EVENTOS DE SEQUÍA.....	35
2.13.1	Periodo de retorno univariado de sequías.	35
2.13.2	Periodo de retorno bivariado de sequías.	36
2.14	PRUEBAS ESTADÍSTICAS	37
2.14.1	Prueba de Ljung-Box.	37
2.14.2	El error cuadrático medio (RMSE)	37
III.	MATERIALES Y MÉTODOS.....	38
3.1	DESCRIPCIÓN DEL ÁREA DE ESTUDIO.....	38
3.1.1	Ubicación Geográfica.....	38
3.1.2	Climatología.....	39
3.2	MATERIALES	39
3.2.1	Programas de cómputo utilizados.	41
3.3	METODOLOGÍA	41
IV.	RESULTADOS Y DISCUSIÓN.....	45
4.1	ANÁLISIS DE LA HOMOGENEIDAD Y TENDENCIA	45
4.2	DETERMINACIÓN DE REGIONES HOMOGÉNEAS	45
4.3	HOMOGENEIZACIÓN DE DATOS POR ZONAS DE ESTUDIO	49
4.4	GENERACIÓN DE SERIES SINTÉTICAS	50
4.4.1	Estandarización de las series.....	50
4.4.2	Cálculo del orden del modelo ARMA(p,q).....	51
4.4.3	Determinación del orden del modelo ARMA usando el BIC	54
4.4.4	Ajuste del modelo ARMA y generación de series sintéticas	55
4.4.5	Validación del modelo	56
4.4.6	Generación y Validación las series	57
4.5	ÍNDICE ESTANDARIZADO DE PRECIPITACIÓN (SPI).....	58
4.5.1	Características de las sequías	60
4.6	SELECCIÓN DE FUNCIONES PROBABILÍSTICAS (PDF)	62
4.7	ANÁLISIS UNIVARIADO	63

4.8	ANÁLISIS BIVARIADO	65
4.8.1	Análisis de correlación.....	66
4.8.2	Análisis de cópulas.....	67
4.9	PERIODO DE RETORNO BIVARIADO PARA EL SPI-3	69
4.10	PERIODO DE RETORNO BIVARIADO PARA UN SPI-12	72
V.	CONCLUSIONES.....	76
VI.	RECOMENDACIONES.....	79
VII.	REFERENCIAS BIBLIOGRÁFICAS	80
ANEXOS.....		89

ÍNDICE DE TABLAS

Tabla 1: Escalas del SPI.....	11
Tabla 2: Guía para identificar los parámetros de modelos autorregresivos.....	24
Tabla 3: Cópulas arquimedianas.....	32
Tabla 4: Estaciones pluviométricas	40
Tabla 5: Prueba de homogeneidad y tendencia de las series	50
Tabla 6: Ajuste del modelo ARMA.....	56
Tabla 7: Caracterización de las sequías	61
Tabla 8: Parámetros de las FDP, para la zona 2; cuenca media Este.	63
Tabla 9: Ajuste para las cópulas Arquimedianas (duración severidad)	68
Tabla 10: Periodo de retorno univariado y bivariado, SPI-3.	72
Tabla 11; Periodo de retorno univariado y bivariado, SPI-12.	75
Tabla 12: Parámetros de las FDP, para la zona 1; Cuenca Baja.	97
Tabla 13: Parámetros de las FDP, para la zona 1; Cuenca Media Este.	97
Tabla 14: Parámetros de las FDP, para la zona 1; Cuenca Alta.	98
Tabla 15: Parámetros de las FDP, para la zona 1; Cuenca Media Oeste.....	98
Tabla 16; Periodo de retorno univariado y bivariado para la zona 1.....	99
Tabla 17; Periodo de retorno univariado y bivariado para la zona 2.....	99
Tabla 18; Periodo de retorno univariado y bivariado para la zona 3.....	99
Tabla 19; Periodo de retorno univariado y bivariado para la zona 4.....	99

ÍNDICE DE FIGURAS

Figura 1: Porcentaje de desastres naturales. Fuente: WMO (2005).	4
Figura 2: Características de las sequías de acuerdo con la teoría de Run.	8
Figura 3: Flujograma para generar pronósticos de series de tiempo	26
Figura 4: Mapa de ubicación de la cuenca hidrográfica del río Mantaro.	38
Figura 5: Diagrama de flujo de la metodología, Fuente: Elaboración propia	43
Figura 6: Dendrograma de regiones homogéneas, Fuente: elaborado en R.	46
Figura 7: Distribución espacial preliminar de las estaciones agrupadas.	47
Figura 8: Delimitación de regiones homogéneas con el método de Thiessen.....	48
Figura 9: Series de tiempo precipitación.	49
Figura 10: Series estandarizadas.	51
Figura 11: Función de autocorrelación función de autocorrelación parcial	52
Figura 12: Mapa de colores para la matriz de AR y MA, para valores BIC.	55
Figura 13: Validación del modelo.	57
Figura 14: SPI para un intervalo de tiempo de 3 meses.	59
Figura 15: SPI para un intervalo de tiempo de 12 meses	60
Figura 16: Funciones de probabilidad para ISP 3	64
Figura 17: Funciones de probabilidad para SPI-12	65
Figura 18: Grafico de dispersión entre las variables de las sequías.	67
Figura 19: Periodo de retorno bivariado SPI-3.....	71
Figura 20: Periodo de retorno bivariado SPI-12.....	73
Figura 21: Serie de precipitación, zona homogenia Cuenca Baja.	91
Figura 22: Serie estandarizada, zona homogénea Cuenca Baja.	91
Figura 23: Serie generada, zona homogénea Cuenca Baja.....	92
Figura 24: Serie de precipitación, zona homogénea Cuenca Media Este.....	92
Figura 25: Serie estandarizada, zona homogénea Cuenca Media Este.	93
Figura 26: Serie generada, zona homogénea Cuenca Media Este.....	93
Figura 27: Serie de precipitación, zona homogénea Cuenca Alta.	94
Figura 28: Serie estandarizada, zona homogénea Cuenca Alta.....	94
Figura 29: Serie generada, zona homogénea Cuenca Alta.	95
Figura 30: Serie de precipitación, zona homogénea Cuenca Media Oeste.	95
Figura 31: Serie estandarizada, zona homogénea Cuenca Media Oeste	96
Figura 32: Serie generada, zona homogénea Cuenca Media Oeste.....	96

RESUMEN

La sequía es un fenómeno complejo que es caracterizada por pocas variables aleatorias correlacionadas, el cálculo de su ocurrencia basado en el análisis univariado no puede dar cuenta de las correlaciones significativas entre las variables. Por lo que, la investigación tuvo como objetivo caracterizar los eventos de sequías en la cuenca del río Mantaro bajo el enfoque bivariado. Para lo cual se usaron datos de precipitación mensual de 43 estaciones meteorológicas agrupadas en cuatro zonas homogéneas: cuenca Alta, Baja, media Este y media Oeste. La metodología consideró el análisis de la homogeneidad y tendencias de las series, la aplicación del método de Box-Jenkins, para modelos autorregresivos de medias móviles, con la finalidad de generar 500 años de series sintéticas, el cálculo del índice de precipitación estandarizado para sequías a corto plazo (SPI-3) y Largo plazo (SPI-12), identificación de las variables, periodo entre sequías, duración, severidad e intensidad. Para el análisis univariado se ensayaron siete distribuciones, mientras que para el análisis bivariado se probaron tres cópulas, finalmente se calculó el periodo de retorno univariado y bivariado. Como resultado se obtuvo cuatro series sintéticas validadas de precipitación mensual ($RMSE < 27.59$; $R^2 > 72.9 \%$; $r > 85.4 \%$), las funciones de distribución más apropiadas para las cuatro variables en estudio fueron la de Pareto Generalizado, Weibull, Log Normal y Pearson. De igual manera las cópulas más convenientes fueron Clayton y Frank. Para el SPI-3 se encontró que los eventos moderados tienen un periodo de retorno bivariado disyuntivo menor de 10 años en la cuenca Baja, cuenca media Este y cuenca media Oeste, sin embargo, para la cuenca Alta los eventos moderados tienen un periodo de retorno menor de 20 años, concluyendo que los eventos de sequía moderados son más ocurrentes en la Cuenca Baja y Media que en la Cuenca Alta. Así mismo, para el SPI-12, se encontró que los periodos de retorno bivariado disyuntivo de estos eventos tienen una ocurrencia menor de 20 años en la cuenca Baja y cuenca media Este, sin embargo, la ocurrencia de estos eventos es mayor a 20 años para la cuenca media Oeste y cuenca Alta, sugiriendo que los eventos severos tienen más probabilidad de ocurrencia en la cuenca Baja y media Este. Mientras que para los eventos extremos no existe diferencias.

Palabras Clave: sequía, SPI, ARMA, bivariado, Mantaro

ABSTRACT

Drought is a complex phenomenon that is characterized by few correlated random variables; the calculation of its occurrence based on univariate analysis cannot account for significant correlations between the variables. Therefore, the objective of the research was to characterize the drought events in the Mantaro river basin under the bivariate approach. For which monthly precipitation data from 43 meteorological stations grouped in four homogeneous zones were used: Upper, Lower, Middle East and Middle West basin. The methodology considered the analysis of the homogeneity and trends of the series, the application of the Box-Jenkins method, for autoregressive models of moving averages, in order to generate 500 years of synthetic series, the calculation of the standardized precipitation index for droughts short-term (SPI-3) and Long-term (SPI-12), identification of variables, the period between droughts, duration, severity and intensity. Univariate and bivariate analysis and finally the calculation of the univariate and bivariate return period. As a result, four validated synthetic series of monthly precipitation were obtained (RMSE <27.59; R²> 72.9%; r> 85.4%), the most appropriate distribution functions for the four variables under study were Generalized Pareto, Weibull, Log Normal and Pearson. Similarly, the most suitable copula was Clayton and Frank. For the SPI-3 it was found that moderate events have a disjunctive bivariate return period of less than 10 years in the Lower Basin, Middle East Basin and Middle West Basin, however, for the Upper Basin moderate events have a return period under 20 years of age, concluding that moderate drought events are more prevalent in the Lower and Middle Basin than in the Upper Basin. Likewise, for the SPI-12, it was found that the disjunctive bivariate return periods of these events have an occurrence of less than 20 years in the Lower Basin and the Middle East Basin, however, the occurrence of these events is greater than 20 years for the Midwest basin and the upper basin, suggesting that severe events are more likely to occur in the lower and middle eastern basin. While for extreme events there are no differences.

Key words: drought, SPI, ARMA, bivariate, Mantaro.

I. INTRODUCCIÓN

La sequía es un período dominado por un clima anormalmente seco, que dura lo suficiente como para producir un desequilibrio grave en el ciclo del agua. Este periodo puede durar varias semanas, meses, incluso años. Puede afectar áreas extensas ocasionando un impacto devastador en la agricultura, en los recursos hídricos, en el medioambiente y en la vida humana, lo que lo vuelve el desastre natural más costoso y extendido (Edwards 1997).

La cuenca del río Mantaro está ubicada en los Andes centrales del Perú y está expuesta a los eventos de sequía. Siendo el último evento registrado en el año 2016 que juntamente con episodios de heladas causaron alrededor de 10 mil hectáreas dañadas y 55 millones de soles en pérdidas en el sector agrícola. Por la naturaleza devastadora de las sequías y por la importancia de la cuenca del Mantaro, es esencial entender este fenómeno y describir sus características, tal como el tiempo entre sequías, duración, severidad e intensidad. Así como también estudiar su ocurrencia, el periodo de retorno univariado y bivariado.

A nivel mundial se han realizado numerosos trabajos de investigación sobre la caracterización de las sequías (Yevjevich 1967; Shiau y Shen 2001; Mishra y Singh 2010). Así mismo existen varios trabajos de investigación realizados en el Perú, uno de ellos fue realizado por Ayala (2017), con el objetivo de comprender el comportamiento de las sequías en la cuenca del Mantaro. Otro trabajo fue reportado por Silva et al. (2008), quien caracterizó los eventos de sequías identificando su duración, severidad e intensidad, bajo un escenario de cambio climático. Sin embargo, estos trabajos analizan la ocurrencia de los eventos de sequías en función de una variable, mientras que la sequía es un fenómeno complejo que se caracteriza por algunas variables correlacionadas, por lo que el análisis univariado de las características de la sequía no es suficiente para explicar las correlaciones significativas entre las variables. Siendo la relación de mayor importancia entre la

duración y severidad (Mirabbasi et al. 2012). Por lo expuesto, es necesario analizar las sequías en la cuenca del Mantaro bajo un enfoque bivariado con el objetivo de entender mejor la ocurrencia de estos eventos

Las cópulas Arquimedianas son técnicas usadas para el análisis de datos bivariado. Estas surgieron como método para abordar problemas multivariados en varias disciplinas y se han constituido como herramienta robusta para cuantificar la estructura de dependencia entre variables correlacionadas en el campo de los recursos hídricos (Chen et al. 2015). Estas cópulas fueron aplicadas al análisis multivariado de las sequías por Shiau (2006), Shiau y Modarres (2009), Mirabbasi et al. (2012), Chen y Montaseri et al. (2018), Tirivarombo et al. (2018) y Guo (2019), con resultados satisfactorios.

Las cópulas Arquimedianas se ajustan a las características bivariadas de los eventos de sequía, las que son derivadas de los registros históricos; sin embargo, estos registros por lo general son insuficientes en extensión para analizar las series de manera consistente. En consecuencia, si se quiere obtener un análisis bivariado confiable de la sequía, se debe contar con el número de datos adecuados (Yoo et al. 2013). La solución a la falta de datos, es la generación de datos sintéticos con las propiedades estadísticas de los datos originales, lo que se puede realizar mediante el uso de modelos estocásticos, también llamados autorregresivos de medias móviles, siendo comúnmente usados en recursos hídricos (Box y Jenkins 1976; Salas et al. 1980). Con los datos generados se proyecta un conjunto de eventos de sequía alternativos que puedan ocurrir, permitiendo el análisis exhaustivo del comportamiento estadístico bivariado de las características de las sequías y también el cálculo del periodo de retorno bivariado. Análisis que son usados en la evaluación de confiabilidad en la gestión y diseño de sistemas de recursos hídricos.

1.1 OBJETIVOS DE LA INVESTIGACIÓN

1.1.1 Objetivos Generales

- Caracterizar los eventos de sequías en la cuenca del río Mantaro bajo el enfoque bivariado.

1.1.2 Objetivos específicos

- Determinar las zonas homogéneas y las series de precipitación representativas en cada zona.
- Generar 500 años de datos de precipitación mensual aplicando modelos autorregresivos de medias móviles.
- Analizar la correlación entre las variables de la sequía.
- Análisis del periodo de retorno univariado y bivariado.
- Determinar el periodo de retorno univariado y bivariado para la duración y severidad.

II. REVISIÓN DE LITERATURA

2.1 SEQUÍAS

Los desastres relacionados con el agua como las sequías, inundaciones, huracanes, tifones y tsunamis, causan terribles daños a la vida y la propiedad humana, mucho más que los daños por terremotos. En la Figura 1, se observa que alrededor del 90 % de los peligros naturales están relacionados con el aire, el clima y agua. De estos desastres naturales las sequías representan el 9 % (Sen 1976).



Figura 1: Porcentaje de desastres naturales. Fuente: WMO (2005).

2.2 DEFINICIÓN DE SEQUÍAS

Las sequías se caracterizan por ser progresiva, por lo que su desarrollo es lento y su existencia es prolongada, perdurando a veces durante muchos años o incluso décadas. Las sequías no están confinadas por características topográficas locales o estructuras geológicas; sin embargo, son más extensas en grandes áreas homogéneas. Existen diferencias con respecto a las variables hidrometeorológicas y socioeconómicas involucradas; por lo que, definir sequía resulta muy complejo, además existe diversas definiciones sobre las cuales existen desacuerdos, sin embargo, cuando se debe definir sequía, se distingue dos tipos de definición: conceptual y operacional (Mishgra y Singh 2010).

Definiciones conceptuales: son definiciones enciclopédicas, en las que no se consideran umbrales o referencias, por consiguiente, no se tiene en cuenta la realidad climática de la zona afectada. Un ejemplo de definición conceptual es la propuesta por la Real Academia Española (1999), que define sequía como "Tiempo seco de larga duración".

Definiciones operacionales: estas definiciones están destinadas a orientar de modo específico sobre los aspectos de un evento de sequía, tales como el inicio, severidad y terminación de la sequía. Las preguntas más frecuentes sobre la descripción objetiva de la sequía son:

- (i) ¿Cuándo es el momento de inicio de la sequía?
- (ii) ¿Qué tan severa es la sequía?
- (iii) ¿Cuándo es el tiempo de finalización de la sequía?

Entonces esta definición se encuentra en función al ámbito de estudio hidrológica, agrícola, meteorológica y geográfica. Por lo que, de forma general la sequía se define como la reducción temporal de la lluvia, escorrentía o la humedad del suelo; por lo tanto, en la definición operacional se intenta identificar el inicio, la gravedad y la terminación del episodio de la sequía. Las estimaciones de los impactos potenciales incluyen algunas definiciones operacionales, las definiciones operacionales también se pueden usar para analizar la frecuencia, severidad y duración de la sequía para determinado período histórico. De acuerdo con lo anterior, Wilhite y Glantz (1985) definen sequía como el fenómeno natural que existe cuando la precipitación ha estado significativamente por debajo de los niveles normales registrados, causando graves desequilibrios hidrológicos que afectan negativamente a la tierra.

2.3 TIPOS DE SEQUÍAS

De acuerdo con Wilhite y Glantz (1985) las sequías son clasificadas en cuatro grupos, que se exponen a continuación.

2.3.1 Sequía meteorológica

Sen (1976) define sequía meteorológica, basada en la cantidad y la duración de la precipitación, como "el período de más de un número particular de días con precipitación menor que una cantidad especificada". Por otro lado Chang y Kleopa (1991) definieron sequía meteorológica como el déficit de precipitación con respecto

al valor promedio, por lo que, en el análisis de este tipo de sequía se usan datos de precipitación diarios, mensuales o anuales. En general, la sequía meteorológica está en función de las precipitaciones y comienza con la disminución de la cantidad de lluvia por debajo del promedio de precipitación, manteniéndose esta condición por periodo prolongado.

2.3.2 Sequía hidrológica

La sequía hidrológica está relacionada con la disponibilidad de agua, estas acontecen cada vez que ocurre una reducción notable de este recurso; por lo que, se estima en los niveles de flujo de agua natural o de agua subterránea, así como también en el agotamiento de almacenamientos de agua en presas y lagos destinados para el suministro de este recurso. La sequía hidrológica no está vinculada directamente a la disminución de la precipitación, por ello acontece en períodos secos cuando la demanda de agua no se puede satisfacer suficientemente como resultado de una condición de clima seco continuo, por consiguiente, el estudio de las sequías hidrológicas es importante en áreas urbanas o regiones industrializadas, así como también en las actividades agrícolas (Sirdaş y Sen 2003).

2.3.3 Sequía agrícola

La sequía agrícola se define en función de la disponibilidad de agua para los cultivos, que permite a las plantas desarrollarse apropiadamente. La fuente de esta humedad es la precipitación, por lo que está en función de la capacidad de infiltración del suelo, esta capacidad varía de acuerdo con las condiciones de humedad, pendiente, tipo de suelo y la intensidad del evento de precipitación (Sen 1976). Por lo tanto la sequía agrícola acontece, cuando la humedad del suelo es insuficiente para mantener el crecimiento y el rendimiento promedio de los cultivos, por consiguiente, la evaluación directa de la sequía agrícola se realiza monitoreo la humedad del suelo en los suelos agrícolas, también se usan índices como el de la Severidad de Sequía de Palmer (Palmer 1965).

2.3.4 Sequía socioeconómica

La sequía socioeconómica ocurre cuando el agua disminuye hasta el punto de generar perjuicios económicos o sociales sobre una población, el cual es producto de la escasez de precipitación. Se diferencia notablemente de los demás tipos de sequía; debido a

que está en función de la relación entre la oferta y la demanda de un bien económico como el agua, los productos agrícolas y la energía hidroeléctrica. La oferta o disponibilidad hídrica varía temporal y espacialmente, debido a su relación directa con la precipitación, a diferencia de la demanda, que fluctúa debido a otros factores, tales como el aumento de la población, el desarrollo de las ciudades, las políticas aplicadas en determinadas regiones (Organización Meteorológica Mundial 2006).

2.4 ANÁLISIS ESTOCÁSTICO DE SEQUÍAS

Şenaut (2015) menciona que las series de tiempo hidrometeorológicas son de naturaleza estocástica, por ser sucesiones de observaciones aleatorias ordenadas en el tiempo, así como también en el espacio. Las variables derivadas de estas series como periodos húmedos y secos, también se comportan de forma estocástico. Por consiguiente, la sequía, que se caracteriza basándose en su duración, severidad e intensidad, también es de naturaleza estocástica. Por lo que, su probabilidad de ocurrencia es casi imposible de predecir con certeza, debido a esta naturaleza temporal, espacial e impredecible, las sequías deben ser caracterizadas mediante procesos estocásticos (Shiau y Shen 2001).

2.4.1 Análisis e identificación de sequías

2.4.2 Caracterización de sequías

Según Yevjevich (1967) la caracterización de las sequías consiste en definir el tiempo entre sequías (Tes), su duración (D), la severidad (S) e intensidad (I). La Figura 2 muestra un gráfico de una variable de sequía hipotética denotada por X_t , la que se cruza en muchos lugares con el nivel de truncamiento X_0 , este es el elemento más básico para derivar los parámetros de la sequía, pudiendo ser constante o estar en función del tiempo. por otro lado, para delimitar de forma temporal el evento de sequía se usa la teoría de corrida (Run), la cual es el intervalo de la serie de tiempo de la variable X_t , cuando esta toma valores que se encuentran por debajo o por encima del nivel de truncamiento seleccionado. Pudiendo haber Run negativas, cuando se encuentra por debajo del umbral y Run positivas, si se encuentran por encima del umbral.

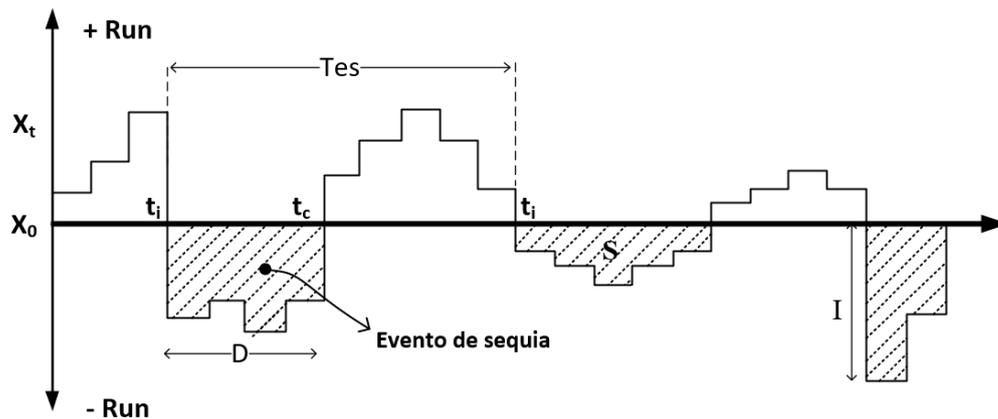


Figura 2: Características de las sequías de acuerdo con la teoría de Run.

Nota. t_i tiempo de inicio de la sequía, t_c tiempo de finalización de la sequía, T_{es} es tiempo entre sequías, D es duración de la sequía, I es intensidad de la sequía, modificado de Mishra y Singh (2010).

Dracup et al. (1980) mencionan que los parámetros estadísticos relacionados a la duración, la magnitud y la intensidad de un evento de sequía tienen los siguientes componentes principales que incluyen.

- (i) Tiempo de inicio de la sequía (t_i): inicio del período de escasez de agua que indica el comienzo de la sequía.
- (ii) Tiempo de finalización de la sequía (t_c): momento en que la escasez de agua se vuelve lo suficientemente pequeña para que las condiciones de sequía ya no persistan.
- (iii) Duración de la sequía (D): periodo durante el cual, el parámetro de sequía está continuamente por debajo del nivel crítico. Es el período de tiempo comprendido entre el inicio y la finalización de una sequía, se expresa en años, meses o semanas, etc.
- (iv) Severidad de la sequía (S): suma acumulada de los valores del parámetro de sequía mientras se encuentra por debajo del nivel crítico.
- (v) Intensidad de la sequía (I): valor mínimo que alcanza el parámetro de sequía en una corrida.

La teoría de la corrida ha sido aplicada para modelar y analizar las sequías por Sen (1976), Dracup et al. (1980), Loaiciga y Leipnik (1996), Mishra y Desai (2005) y Panu y Sharma (2009).

2.5 ÍNDICES DE SEQUÍAS

Basándose en los cuatro tipos de sequías expuestas en la sección anterior, existen más de 150 índices de sequía propuestas, los cuales son ampliamente aceptados como medidas para monitorear la variabilidad espacial y temporal de la escasez de agua (Zargar et al. 2011). El uso de estos índices de sequía depende de la región climática a evaluar, el tipo de sequía y el propósito del estudio. Mientras que su cálculo de estos índices se realiza basándose en información como precipitación, evaporación potencial, humedad del suelo, caudales, nivel de aguas subterráneas, escasez de suministros, etc. (AMS 1997).

2.5.1 Índice de Precipitación Estandarizado

McKee et al. (1993) propusieron el índice de precipitación estandarizada (SPI, por sus siglas en inglés *Standardized Precipitation Index*), con la finalidad de seguimiento y análisis de la sequía meteorológica. Este índice se determina con la diferencia de precipitación de la media para un período de tiempo específico dividido por la desviación estándar, donde la media y la desviación estándar se determinan a partir de registros históricos de precipitación diarios, mensuales o anuales. Sin embargo, para la precipitación acumulación de 12 meses o menos, al no tener la distribución normal, en la mayoría de los casos, se realiza la transformación de los datos, el cálculo resultante, es una serie estandarizada linealmente proporcional al déficit de precipitación, permitiendo especificar la probabilidad, el porcentaje del promedio y el déficit acumulado. Los valores del SPI representan la probabilidad de ocurrencia, independientemente del período de tiempo, la ubicación y el clima. Por lo que, la Organización Meteorológica Mundial recomienda como índice estándar de seguimiento de las sequías.

2.5.2 Cálculo de ISP

La caracterización de un evento de sequía se efectúa identificando el tiempo de inicio y finalización, duración, severidad e intensidad, estas variables pueden calcularse a partir del valor del Índice de Precipitación Estandarizado, finalmente el SPI se determina sobre la serie de datos de precipitación, siguiendo los pasos propuesto por McKee et al. (1993), que a continuación se detallan.

a) Ajuste a la serie de datos mensuales a la función gamma

$$g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \text{ para } x > 0$$

Donde:

$\alpha > 0$ es el parámetro de forma

$\beta > 0$ es el parámetro de escala

$\Gamma(\alpha)$ es la función gamma y está definido

$$\Gamma(\alpha) = \lim_{n \rightarrow \infty} \prod_{y=1}^n \frac{n! n^{y-1}}{y + \alpha} \equiv \int_0^\infty y^{\alpha-1} e^{-y} dy$$

Para ajustar los datos de precipitación a la distribución gamma se requiere determinar los parámetros de forma y escala. Edwards (1997) sugiere utilizar las siguientes aproximaciones para calcular los parámetros de la función gamma.

$$\alpha = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right) \quad \beta = \frac{\bar{x}}{\alpha}$$

Donde:

$$A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n}$$

Dado que la función gamma no está definida para valores de $x = 0$, sin embargo, un registro de precipitación puede contener valores iguales a cero, por lo que, la probabilidad acumulada se convierte en:

$$H(x) = q + (1-q)G(x)$$

Donde: q Es la probabilidad.

La probabilidad acumulada $H(x)$ se transforma luego en la variable aleatoria normal estándar Z con media cero y varianza uno, el cual es el valor del SPI. Se emplea la conversión aproximada proporcionada por Abramowitz y Stegun (1965) como alternativa:

$$Z = SPI = - \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \quad \text{para } 0 < H(x) \leq 0.5$$

$$Z = SPI = + \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \quad \text{para } 0.5 < H(x) < 1$$

Donde:

$$t = \sqrt{\ln \left(\frac{1}{(H(x))^2} \right)} \quad \text{para } 0 < H(x) \leq 0.5$$

$$t = \sqrt{\ln \left(\frac{1}{(1-H(x))^2} \right)} \quad \text{para } 0.5 < H(x) < 1$$

Además

$$c_0 = 2.515517, c_1 = 0.802853, c_2 = 0.010328, d_1 = 1.432788, d_2 = 0.189269, d_3 = 0.001308$$

La principal ventaja de este método es su estandarización, de modo que los valores representan la misma probabilidad de ocurrencia independientemente del período de tiempo, la ubicación y clima. Los intervalos categóricos iguales tienen diferentes probabilidades de ocurrencia. En la Tabla 1 se muestra la diferencia de probabilidad (ΔP) para cada intervalo de SPI. La ΔP para el intervalo SPI de -1.0 y -1.5 es 9.2 %, los que corresponde a un evento de sequía moderada, la ΔP para el intervalo SPI de -1.5 y -2.0 es 4.4 % correspondiendo a un evento de sequía severa. El SPI ha sido diseñado para comprender las reducciones de precipitación en múltiples escalas de tiempo como diarios, mensuales, anuales.

Tabla 1: Escalas del SPI

Orden	Clase	Valores de Índice	Probabilidad	ΔP
1	Extremadamente húmedo	$SPI \geq 2.0$	0.977-1.000	0.023
2	Muy húmedo	$1.5 \leq SPI < 2.0$	0.933-0.977	0.044
3	Moderadamente húmedo	$1 \leq SPI < 1.5$	0.841-0.933	0.092
4	Casi normal	$-1 \leq SPI < 1.0$	0.159-0.841	0.682
5	Sequía moderada	$-1.5 < SPI \leq -1.0$	0.067-0.159	0.092
6	Sequía severa	$-2 < SPI \leq -1.5$	0.023-0.067	0.044
7	Sequía extrema	$SPI < -2.00$	0.000-0.023	0.023

Fuente: McKee et al. (1993).

2.6 CARACTERIZACIÓN DE LAS SEQUÍAS

Tiempo entre sequías (T_{es}): es el intervalo entre el inicio de un evento y el inicio del siguiente evento, por lo que se considera el intervalo de tiempo entre el primer valor SPI menor a -1 de un determinado evento de sequía, hasta el primer valor SPI menor a -1 que da inicio a la siguiente sequía, se define con la siguiente fórmula.

$$T_{es} = T_{i+1} - T_i$$

Donde: T_{es} representa al tiempo entre sequías. T_i es el tiempo de inicio de un evento cualquiera y T_{i+1} es el tiempo de inicio del evento inmediatamente después.

Duración (D): Es el intervalo de tiempo comprendido entre el inicio y fin de un evento. El tiempo de inicio se considera desde que el SPI alcanza un valor menor igual a -1. Mientras que, el tiempo de finalización del evento se considera cuando SPI alcanza un valor mayor a -1, lo que se puede expresar de la siguiente forma.

$$D = T_i - T_c$$

Donde: D representa la duración de la sequía. T_i es el tiempo de inicio de la sequía y T_c es el tiempo de finalización de la sequía.

Severidad (S): Esta determinado por el valor absoluto de la suma de todos los ISP que pertenecen a un evento de sequía y se calcula con la siguiente formula:

$$S = \left| \sum SPI \right|$$

Intensidad (I): En este trabajo, se define intensidad de una sequía como el valor absoluto del menor valor que tiene el SPI en un evento, se calcula usando la siguiente formula.

$$I = \left| MIN(SPI) \right|$$

2.7 SERIES TEMPORALES Y SU ANÁLISIS

Hipel y McLeod (1994) definen serie de tiempo como el conjunto de observaciones ordenadas cronológicamente, el orden de aparición de las observaciones es importante en su análisis. Ejemplos de series de tiempo son, el registro diario de precipitaciones,

el registro mensual de caudales de un río, etc. Una de las características intrínsecas de una serie temporal es la dependencia adyacente de las observaciones, la naturaleza de esta dependencia entre las observaciones es interés práctico en su análisis. Por lo que el análisis de las series temporales se basa en técnicas que estudian la dependencia adyacente de las observaciones, estos análisis requieren el desarrollo de modelos estocásticos y dinámicos. Así mismo Box et al. (2015) mencionan que una de las finalidades de estos análisis es de pronosticar las condiciones climáticas, basándose en la secuencia de tiempo que evolucionaron en los días anteriores.

Salas et al. (1980), definen serie de tiempo, de N sucesivas observaciones aleatorias (x_1, x_2, \dots, x_N), donde x es una variable aleatoria cuyos subíndices representan intervalos iguales de tiempo o espacio. La serie definida está constituida por magnitudes aleatorias que evolucionan en el tiempo o espacio, por lo que se denomina serie estocástica.

2.7.1 Test de homogeneidad o consistencia de las series de tiempo

Para el cálculo del SPI se requiere registros de precipitación homogéneos y consistentes; sin embargo, las series de tiempo comúnmente no son homogéneas, ni consistentes, debido a factores naturales o procesos antropogénicos. Las series no homogéneas tienen errores sistemáticos, los que afectan varias características de las series tal como la media, desviación estándar, correlación serial, las tendencias, datos atípicos. Por consiguiente, el primer análisis que se debe realizar es la identificación, descripción y remoción de la no homogeneidad (Salas et al. 1980).

Existen varios métodos para detectar la no homogeneidad en las series. Como el método gráfico, el cual se utiliza a menudo en climatología e hidrología para tener una idea preliminar del comportamiento de la serie en el tiempo. Otro método, es la curva de doble masa que se obtiene trazando los montos acumulados de la estación considerada confiable contra los montos acumulados de un conjunto de estaciones vecinas, los puntos trazados tienden a caer a lo largo de una línea recta en condiciones de homogeneidad, de modo contrario existirán quiebres en la recta. En lugar de la curva de doble masa, también se pueden trazar las desviaciones acumuladas de algún valor promedio, las desviaciones acumuladas tienen ventaja de que los cambios en la

cantidad media de lluvia son más fáciles de reconocer, la gráfica de las desviaciones acumulativas a veces se denomina curva de masa residual (Buishand 1982).

La homogeneidad implica que todos los datos recopilados pertenezcan a la misma población estadística, con la media invariante en el tiempo. Por lo que, las pruebas estadísticas para verificar la homogeneidad de una serie, se basan en la evaluación del cambio en el valor medio. Existen cuatro pruebas de homogeneidad comúnmente usadas en análisis hidrológico: ratio Von Neumann, Buishand, Standard normal (SNHT) y Pettitt. Siendo estas pruebas complementarias, las pruebas de Buishand, SNHT y Pettitt identifican si hay un salto en la serie de tiempo y determinan el punto de ruptura, mientras que la prueba de Von Neumann, supone que la serie no se dispersa aleatoriamente y no proporciona información sobre cuándo se deteriora la homogeneidad. Si bien la prueba SNHT es sensible a las rupturas al principio y al final de la serie temporal, la prueba de Pettitt y la prueba de Buishand son más susceptibles a detectar roturas en la mitad de la serie (Arikan et al. 2018).

2.7.2 Ratio de Von Neumann

Esta prueba paramétrica de homogeneidad tiene como hipótesis nula (H_0) que los datos son homogéneos, mientras que hipótesis alterna indica que existe al menos un punto en la que hay un cambio en los datos. La prueba de ratio de Von Neumann no determina un lugar específico donde la homogeneidad se ve afectada y no proporciona información sobre cuándo se deteriora la homogeneidad, por lo que esta prueba es complementaria a otras (Wijngaard et al. 2003). En esta prueba N describe la proporción cuadrática media de la variación de la varianza en sucesión expresada mediante la siguiente ecuación.

$$N = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x}_{i-1})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Donde: \bar{x} , es la media de los valores observados. Si el valor N calculado supera el valor N crítico, la hipótesis H_0 se acepta y la serie de tiempo se considera homogénea.

2.7.3 Test de Buishand

En esta prueba se asume que los datos son independientes, aleatorios y se distribuyen normalmente. Esta prueba es sensible a las rupturas en la mitad de la serie (Bickici y

Kahya 2018). Para su cálculo se utiliza las sumas parciales ajustadas que se definen como:

$$S_0^* = 0; \quad S_k^* = \sum_{i=1}^k (x_i - \bar{x}), \quad k = 1, 2, \dots, N$$

Donde: \bar{x} , es la media de los valores observados. K , es el número de observaciones en las que se ha producido un punto de ruptura. Las sumas parciales ajustadas redimensionadas se obtienen dividiendo S_k^* por la desviación estándar de la muestra.

$$S_0^{**} = \frac{S_k^*}{D_x}, \quad k = 1, 2, \dots, N$$

$$D_x = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

El estadístico que se usa para analizar la homogeneidad es expresado como:

$$Q = \max_{0 \leq k \leq N} |S_k^{**}|$$

$\frac{Q}{\sqrt{N}}$ el valor se compara con el valor crítico dado por

$$x_{k\alpha} = \left(\frac{\ln(n^3 + n^2)}{6} \right)^{\frac{1}{2}}$$

Si el valor anterior no excede al valor crítico, se acepta la hipótesis nula, concluyendo que la serie es homogénea.

2.7.4 Prueba estándar de homogeneidad normal

También conocida por SNHT, por sus siglas en inglés. Es una prueba de homogeneidad que se usa a menudo en estudios climáticos. Este método, se ha utilizado con éxito en pruebas de muchas escalas climáticas e hidrológicas, es flexible y fácil de usar (Arikan et al. 2018). Dado una serie, con un quiebre ubicado en el punto K , $T(k)$ alcanza su valor máximo en $k = K$. El estadístico de prueba T_0 se define como:

$$T_0 = \max(T(k)) \quad \text{para } 1 \leq k < n$$

Donde:

$$T(k) = k\bar{z}_1^2 + (n-k)\bar{z}_2^2; \quad k = 1, 2, \dots, n$$

$$\bar{z}_1 = \frac{1}{k} \frac{\sum_{i=1}^k (x_i - \bar{x})}{s} \quad y \quad \bar{z}_2 = \frac{1}{n-k} \frac{\sum_{i=k+1}^n (x_i - \bar{x})}{s}$$

La fórmula para el estadístico es el siguiente:

$$T_0 = \frac{n(T(n))^2}{n-2 + (T(n))^2}$$

Si T_0 excede el valor del estadístico de prueba, se rechaza la hipótesis nula.

2.7.5 Prueba de Pettitt

Este método no paramétrico propuesto por Pettitt (1979) se usa para detectar el punto de cambio en una serie de tiempo, identificando el punto de cambio en una escala mensual o anual. La hipótesis nula, indica una distribución independiente y aleatoria de la serie, la hipótesis alternativa indica la existencia de un cambio repentino. Esta prueba es de rango no paramétrico. Para calcular las estadísticas, se utilizan los rangos $r_1 \dots r_n$ de $Y_1 \dots Y_n$:

$$x_k = 2 \sum_{i=1}^k r_i - k(n+1), \quad k = 1 \dots n$$

Si hay una ruptura en el mes o año K , la estadística es máxima o mínima cerca del año $k = K$:

$$x_k = \max_{1 \leq k \leq n} |x_k|$$

La significancia estadística para un nivel probable α es:

$$x_{k\alpha} = \left(\frac{\ln \alpha (n^3 + n^2)}{6} \right)^{\frac{1}{2}}$$

Los valores críticos de esta prueba se dan si los valores X_K no exceden los valores críticos, se acepta la hipótesis H_0 ; Es decir, es homogénea.

2.7.6 Análisis de la estacionariedad

Una serie temporal es estrictamente estacionaria si sus propiedades estadísticas no varían con el tiempo. Es decir, si al seleccionar dos intervalos de tiempo no superpuestos de una serie dada, las dos subseries se verán casi idénticas. De hecho, ambas subseries diferirán visualmente unas de otras, pero se dispersarán alrededor del mismo valor medio. Por lo tanto, una serie temporal estacionaria no puede tener ninguna tendencia o componente periódico.

Existen dos enfoques para verificar la estacionariedad de las series de tiempo: la primera, es el enfoque paramétrico generalmente se utiliza en el dominio del tiempo; la segunda, es el enfoque no paramétrico el cual es comúnmente utilizado en el dominio de la frecuencia. En hidrología, se utilizan ambos enfoques. Las pruebas no paramétricas, no se basan en la suposición de que la población se distribuye normalmente, en consecuencia, estas pruebas se aplican más que las pruebas paramétricas. Sin embargo, las pruebas no paramétricas son menos robustas que las pruebas paramétricas, para llegar a la misma conclusión con el mismo nivel de confianza, las pruebas no paramétricas requieren entre 5 a 35 por ciento más de datos que las pruebas paramétricas (Bethea 1995).

2.7.7 Análisis de tendencias Mann - Kendall

Zubieta et al. (2017) menciona que la prueba no paramétrica Mann–Kendall es adecuada para el análisis de tendencias en datos de precipitación anual, Yue et al. (2002) aplica este método con buenos resultados a series hidrológicas para detectar tendencias. Para aplicar este método se propone el siguiente estadístico.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i)$$

Donde x_i es el dato tiempo actual para la serie de tiempo $i = 1, 2, \dots$

$$\text{Sgn} \begin{cases} +1 & > (x_j - x_i) \\ 0 & = (x_j - x_i) \\ -1 & < (x_j - x_i) \end{cases}$$

Cuando número de datos ($n \geq 10$) el estadístico S sigue la distribución normal con

$$\text{media } E(S) = 0 \text{ y la varianza } \text{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^m t(t-1)(2t+5)}{18}$$

Donde t_i son las colas de la muestra de la serie temporal. La prueba estadística de Z_c es como sigue:

$$z_c = \begin{cases} \frac{s-1}{\sqrt{\text{Var}(S)}} & S > 0 \\ 0 & S = 0 \\ \frac{s-1}{\sqrt{\text{Var}(S)}} & S < 0 \end{cases}$$

Donde: Z_c sigue una distribución normal, el valor positivo representa una tendencia positiva, mientras que el valor negativo representa una tendencia descendente para el periodo evaluado. Si el nivel de significancia de la prueba es $(\alpha, Z_c \geq Z_{\alpha/2})$ entonces la tendencia de los datos es considerada significativa. El test de Mann Kendall prueba la hipótesis nula de no tendencia, la hipótesis alterna indica que existe tendencia en los datos, sin embargo no indica si esta es positiva o negativa (Mann, 1945; Kendall, 1975).

Escriba el texto aquí

Hipel y McLeod (1994) plantearon una extensión multivariada de la prueba estadística Mann – Kendall para detectar tendencias en series estacionales, asumiendo una serie de tiempo completa X , de n años con m estaciones por año. Tal que X está dado por:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

La hipótesis nula (H_0) de esta prueba plantea que para cada m estación de n observaciones son independientes e idénticamente distribuidos, mientras que la hipótesis alterna es, existe tendencia monótona de la serie. La matriz de rangos es denotada como sigue.

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{22} & \cdots & R_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ R_{n1} & R_{n2} & \cdots & R_{nm} \end{bmatrix}$$

Donde las n observaciones de cada temporada o columna en R se clasifican entre sí. Por lo tanto, el rango de x_{jg} , que es el j -ésimo dato en la g -ésima estación, está definida por:

$$R_{jg} = \left[n + 1 + \sum_{i=1}^n \text{sgn}(x_{jg} - x_{ig}) \right] / 2$$

Y cada columna de R es una permutación de $(0, 1, 2, \dots, n)$, la prueba estadística de Mann – Kendall para la g -ésima estación es:

$$S_g = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_{jg} - x_{ig}), \quad g = 1, 2, \dots, m$$

Donde: S_g tiene una distribución normal asintótica

$$E[S_g] = 0$$

Prueba no paramétrica

$$\text{Var}[S_g] = \sigma_g^2 n(n-1)(2n+5)/18$$

Donde el estadístico de Kendall para el g -ésima estación está definido como:

$$\tau_g = \frac{S_g}{\frac{1}{2}n(n-1)} = \frac{S_g}{\binom{n}{2}}$$

Debido a que τ_g es múltiplo de S_g , la distribución de τ_g puede obtenerse de la distribución de S_g . En particular, τ_g es normalmente distribuido asintóticamente, donde:

$$E[\tau_g] = 0 \quad E[\tau_g] = \frac{1}{\left(\frac{1}{2}n(n-1)\right)^2} \left(\frac{n(n-1)(2n+5)}{18} \right)$$

Aritméticamente es más conveniente tratar con S_g que con τ_g . Por lo tanto, la prueba estadística Mann-Kendall es:

$$S' = \sum_{g=1}^m S_g$$

2.8 GENERACIÓN DE SERIES SINTÉTICAS

Hipel y McLeod (1994) informan que en el diseño, planificación y operación de los sistemas de recursos hídricos a menudo se necesitan buenas estimaciones del comportamiento futuro de las variables hidrológicas. Por ejemplo, en la generación de energía en hidroeléctricas, en la disponibilidad de agua en los sectores agrícolas o para el abastecimiento de agua potable en poblaciones. Para ello se requieren pronósticos de los flujos proyectados para los próximos períodos de tiempo. En este sentido, el objetivo del pronóstico es proporcionar predicciones precisas de lo que sucederá en el futuro, para la gestión adecuada de los recursos hídricos. La generación de series sintéticas también es usado en el análisis de sequías, debido a la poca disponibilidad de datos en los registros hidrometeorológicos utilizados (Tosunoğlu y Onof 2017), uno de estos métodos para generar series sintéticas son los modelos autorregresivos propuestos por Box y Jenkins (1976).

2.8.1 Procesos autorregresivos (AR)

Las series de tiempo en hidrología consisten en puntos de datos que son serialmente dependientes. Por lo que se estima un coeficiente o conjuntos de coeficientes que describe cada punto de datos consecutivos de la serie a partir de puntos de datos pasados. Esta dependencia puede resumirse mediante la siguiente expresión de proceso autorregresivo.

$$x_t = \xi + \phi_1 x_{(t-1)} + \phi_2 x_{(t-2)} + \phi_3 x_{(t-3)} + \dots + \varepsilon_t$$

Donde:

- $x_t =$: datos de la variable x en el tiempo t.
 $x_{(t-1)}, x_{(t-2)}, x_{(t-3)}$: variable de x en los tiempos previos $t-1, t-2, t-3$
 ξ : constante (intercepto o media poblacional).
 ϕ_1, ϕ_2, ϕ_3 : parámetros del modelo autorregresivo.
 ε_t : ruido blanco.

Para determinar x_t , la función debe estar constituido por un componente aleatorio y una combinación lineal de los datos anteriores, la ecuación representa el proceso autorregresivo infinito. Sin embargo, en la práctica la media poblacional se reemplaza

por la media de la muestra y el orden autorregresivo es reducido a p . de esta forma la ecuación queda reducida como sigue:

$$x_t = \bar{x} + \phi_1 x_{(t-1)} + \phi_2 x_{(t-2)} + \dots + \phi_p x_{(t-p)} + \varepsilon_t$$

El orden del proceso autorregresivo está definido por el mayor valor de p , para cada $\phi_p \neq 0$. Así, para $p = 1$, el proceso autorregresivo AR es de primer orden y para $p = 2$, el proceso es de segundo orden. El primero y segundo orden de estos procesos autorregresivo son denotados por AR(1) y AR(2) respectivamente. En términos generales AR(p) indica el proceso autorregresivo de orden p , (Box $\text{\textcircled{C}}$ y Jenkins $\text{\textcircled{C}}$, 1976).

2.8.2 Procesos de medias móviles

Aparte de la dependencia serial de los datos, como es el caso del proceso autorregresivo, cada uno de los datos puede ser afectado por el error aleatorio pasado, lo que no se considera en el modelo autorregresivo. El error aleatorio pasado, puede ser expresado por el proceso de medias móviles, dado por la ecuación:

$$x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} - \theta_2 \varepsilon_{(t-2)} - \theta_3 \varepsilon_{(t-3)} - \dots$$

Donde:

$x_t =$: datos de la variable x en el tiempo t .

$\varepsilon_{(t-1)}, \varepsilon_{(t-2)}, \varepsilon_{(t-3)}$: componente del error aleatorio de los datos en los tiempos previos $t-1, t-2, t-3$.

μ : constante o media poblacional.

$\theta_1, \theta_2, \theta_3$: parámetros del modelo promedios móviles.

De acuerdo con la ecuación anterior, cada dato de la serie está constituido por un componente de error aleatorio y una combinación lineal de este. La ecuación representa el proceso de medias móviles infinito. Sin embargo, puede ser reescrito reemplazando la media poblacional por la media de la muestra y reduciendo el orden del proceso de medias móviles de infinito a q , tal como se indica en la siguiente formula:

$$x_t = \bar{x} + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} - \theta_2 \varepsilon_{(t-2)} - \dots - \theta_q \varepsilon_{(t-q)}$$

Escriba el texto aq

El orden del proceso de medias móviles, está definido por el mayor valor de q , para $\theta_q \neq 0$, así para $q = 1$ la media móvil es de primer orden y para $q = 2$ el proceso es de segundo orden. El primero y segundo orden del proceso de medias móviles son simplificadas y denotadas por MA(1) y MA(2) respectivamente, el proceso de MA de orden q puede ser denotado por MA(q)

2.8.3 Proceso autorregresivo de medias móviles (ARMA)

Los procesos autorregresivos AR(p) y media móvil MA(q) cuando se consideran en conjunto forman un caso especiales de modelo estocástico, al que se le denomina procesos autorregresivos de media móvil (ARMA) de orden (p, q) denotado por ARMA(p, q). El cual representa un auténtico proceso estocástico de la serie x_t , representada por la siguiente expresión:

$$x_t = \bar{x} + \phi_1 x_{(t-1)} + \phi_2 x_{(t-2)} + \dots + \phi_p x_{(t-p)} + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} - \theta_2 \varepsilon_{(t-2)} - \dots - \theta_q \varepsilon_{(t-q)}$$

El proceso ARMA ($p, 0$) con p mayor e igual a uno es el proceso AR (p), mientras que el proceso ARMA ($0, q$) con q mayor e igual a uno es el proceso MA (q).

2.8.4 Metodología para la aplicación del modelo ARMA

La metodología para aplicar cualquiera de los tres modelos estocásticos (AR, MA y ARMA) a una serie de tiempo fueron propuestas por Box y Jenkins (1976) y Salas et al. (1980). Siendo los siguientes:

a. Identificación del modelo ARMA

la serie de tiempo de entrada para ser simulada por el modelo ARMA debe ser homogénea y estacionaria, con media, varianza y autocorrelación constante en el tiempo. Por lo que, primero, se debe verificar la estacionariedad de la serie, si existe tendencias en la serie se deben remover; segundo, se debe verificar la estacionalidad, si existe patrones estacionales se deben remover. Finalmente, la serie de tiempo se estandariza aplicado la siguiente formula.

$$z_\tau = \frac{y_\tau - \mu_\tau}{\sigma_\tau}$$

Donde: μ_τ y σ_τ es el promedio y la desviación estándar de la serie, obteniéndose una serie estandarizada (Z_τ), con media cero, varianza σ^2 y con distribución normal.

b. Identificación del orden del modelo ARMA

Para determinar el orden del modelo debe definirse los parámetros autorregresivos p y de media móvil q . El orden del modelo se selecciona de tal manera que el modelo sea efectivo y parsimonioso, un modelo parsimonioso tendrá la menor cantidad de parámetros y el mayor número de grados de libertad, entre todos los modelos estocásticos que se ajusten a las series (Hipel y McLeod 1994). La metodología para determinar el orden del modelo inicia analizando los correlogramas que se generan a partir de la función de autocorrelación (ACF, por sus siglas en inglés) y de la función de autocorrelación parcial (PACF, por sus siglas en inglés). Si bien este análisis no es sencillo y requiere no solo de una vasta experiencia, sino también de una buena cantidad de pruebas con modelos estocásticos alternativos, es considerada como esencial y base para subsiguientes análisis (Salas et al. 1980).

La autocorrelación es la correlación de la serie temporal con sus propios puntos de datos pasados, se refiere a la correlación lineal entre los miembros de una serie de números dispuestos en el tiempo, también se denomina correlación rezagada o correlación serial. La autocorrelación positiva puede considerarse como la forma específica de persistencia de los valores pasados en los presentes, es la tendencia de un sistema a permanecer en el mismo estado (Kisiel 1969).

La ACF mide el grado de autodependencia lineal de las series temporales. La autocorrelación c_k entre x_k y x_{t+k} puede ser determinado por:

$$r_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Donde: r_k es usualmente llamando coeficiente de autocorrelación del rezago k , el cual representa el rezago en el tiempo, entre el par correlacionado (x_t, x_{t+k}) , \bar{x} es la media muestral y N el tamaño de muestra, a la colección de valores de r_k para cada valor de $k = 0, 1, 2, \dots$ se le llama ACF.

La PACF está definida por los coeficientes de correlación parcial entre las series de tiempo y los rezagos de las series a lo largo del tiempo. La autocorrelación parcial en

el rezago k es la autocorrelación entre x_t y x_{t-k} que no está explicada por los rezagos 1 a $k - 1$. La autocorrelación parcial del proceso AR (p) es cero en retardo mayor que o igual a $(p + 1)$. El algoritmo detallado con las expresiones matemáticas para calcular la PACF se pueden encontrar en Box y Jenkins (1976), Brockwell y Davis (2002).

Tabla 2: Guía para identificar los parámetros de modelos autorregresivos

Nº	Parámetros	Características de ACF	Características de PACF
1	AR (1)	Decaimiento exponencial	Significativo en el rezago 1, no correlacionado para los otros rezagos.
2	AR (2)	Patrón de decaimiento en forma de onda de seno o decaimiento exponencial	Significativos para los rezagos 1 y 2, no correlacionados para los demás.
3	MA(1)	Significativo para el rezago 1, no correlacionado para los demás rezagos.	Desciende exponencialmente
4	MA(2)	Significativo para el rezago 1 y 2, no correlacionado para los demás rezagos.	Patrón de decaimiento en forma de onda de seno o decaimiento exponencial
5	ARMA(1,1)	Decaimiento exponencial comenzando en el rezago 1	Decaimiento exponencial comenzando en el rezago 1

Nota: ACF es función de autocorrelación, PACF es función de autocorrelación parcial. Modelo autorregresivo de orden uno y dos AR(1) y AR(2). Modelo de medias móviles de orden uno y dos MA(1) y MA(2). Modelo autorregresivo de medias móviles 1,1 ARMA(1,1).

Pankratz (2009) formuló pautas generales para identificar cinco modelos estocásticos básicos basados en la forma y características del autocorrelograma y el autocorrelograma parcial, este criterio se muestra en la Tabla 2. Donde la mayoría de los patrones de series de tiempo se pueden aproximar satisfactoriamente a uno de los cinco modelos básicos. Sin embargo se pueden encontrar más detalles y sugerencias para seleccionar el orden del modelo estocástico en Box y Jenkins (1976), Hipel y McLeod (1994), Machiwal y Jha (2012) y Box et al. (2015).

c. Método alternativo para determinar el orden del modelo ARMA

Otra forma de establecer el orden del modelo ARMA es el basado en la estimación de la varianza de los parámetros del modelo estimado. Para lo cual existen dos métodos; el primero, llamado Criterio de información de Akaike (AIC) (Burnham y Anderson 2004) el cual se define como:

$$AIC(p, q) = \log(\sigma_{p,q}^2) + 2 \frac{p + q + 1}{n + 1}$$

El segundo, propuesto por Burnham y Anderson (2004), llamado método del Criterio de Información Bayesiano (BIC), cuya ecuación es:

$$BIC(p, q) = \log(\sigma_{p,q}^2) + 2 \frac{(p+q)\log(n+1)}{n+1}$$

La estimación de la varianza ($\sigma_{p,q}^2$) para ambos métodos se vuelve pequeña a medida que el orden ($p + q$) aumenta. Por lo tanto, los términos aditivos en los criterios anteriores sirven como penalizaciones para valores grandes de p y q , evitando el ajuste excesivo de los datos seleccionando. No existe razón específica para usar uno u otro modelo, sin embargo, se debe tener en cuenta que AIC tiene la tendencia a no subestimar el orden del modelo, mientras que BIC se prefiere para muestras grandes.

d. Generación de series sintéticas

Para generar las series sintéticas usando el modelo ARMA, primero se selecciona el orden del modelo, luego se determina sus parámetros y finalmente se valida el modelo, entonces el modelo se encuentra apto para ser usado. Por lo general, los pronósticos se hacen para el futuro, de modo que estos nuevos valores calculados estén más allá de los puntos de datos incluidos en la serie de entrada. Si el proceso de estimación se realiza en una serie transformada o diferenciada, esta debe integrarse antes de que se generen los pronósticos. La integración es el proceso inverso de la diferenciación (Machiwal y Jha 2012).

e. Pronóstico de datos de series de precipitación usando el modelo ARMA

Palit y Popovic (2005) sugieren que una vez que se ha construido el modelo ARMA(p,q), este se puede usar para pronosticar los valores futuros de la serie. Así mismo, Hipel y McLeod (1994) mencionan que los pronósticos se calculan después de que el modelo de serie de tiempo más apropiado se ajusta a una secuencia de datos observados. Estos pronósticos se realizan para uno o más pasos de tiempo en el futuro, sin embargo, estos modelos también se utilizan en la simulación de las series. En la Figura 3 se muestra el procedimiento general para obtener pronósticos.

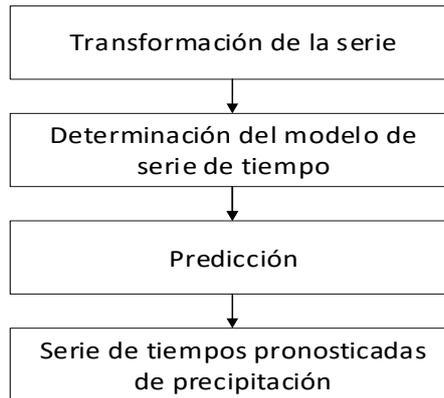


Figura 3: Flujograma para generar pronósticos de series de tiempo

z_t representa el último valor conocido de la serie de tiempo no transformada, observada en el tiempo t , por otro lado, se tiene el modelo ARMA que se ajusta a la serie histórica hasta el tiempo t . Para pronosticar valores futuros de la serie en el tiempo $t + 1$, se utiliza el modelo ARMA y las observaciones más recientes de la serie, obteniéndose datos que vendrían a ser las predicciones z_{t+1} , denotando en general dichos valores por $\hat{z}_t(l)$. En la serie el tiempo t se conoce como el tiempo de origen para el pronóstico, mientras que el tiempo denotado por l , es el tiempo de pronóstico. Los valores pronosticados, para los tiempos: $l = 1, 2$ y 3 , se denotan por $\hat{z}_t(1)$, $\hat{z}_t(2)$ y $\hat{z}_t(3)$ respectivamente.

2.9 ANÁLISIS UNIVARIADO DE SEQUÍAS

El análisis univariado en hidrología es importante para la estimación de la ocurrencia de eventos de sequías. Por lo que, mediante este análisis se puede estimar la probabilidad de ocurrencia de un determinado evento. Así mismo se puede determinar el intervalo de ocurrencia de la severidad e intensidad de las sequías (Sharma 1997). El procedimiento para el análisis de frecuencia en hidrología incluye principalmente dos pasos; primero, selección de una función de distribución teórica apropiada; segundo, estimación de los parámetros de la distribución. Existe un gran número de funciones teóricas, sin embargo, las más utilizadas en el ámbito de los Recursos hídricos son la función Exponencial, Gumbel, Weibull, Gamma y Log-normal y distribuciones de tres parámetros, como valor extremo general y Pearson tipo III. Para

seleccionar una función teórica, se debe determinar si la función se ajusta a los datos observados apropiadamente, para lo cual, se usan pruebas estadísticas como la de Kolmogorov-Smirnov, Anderson-Darling y Chi cuadrado (Pankratz 2009).

2.10 ESTIMACIÓN DE PARÁMETROS

Los métodos más usados para estimar los parámetros de una función, son el método de momentos (MOM) y el método de máxima verosimilitud (ML). MOM se caracteriza por su simplicidad en el proceso de obtención de los valores de los parámetros, sin embargo, una de sus limitaciones es la influencia de los datos atípicos sobre los momentos de la serie. Por otro lado, se tiene el método de ML el cual se caracteriza por tener un procedimiento más complejo y obtener estimaciones más precisas. Sin embargo, en algunos casos el MOM puede ayudar a la resolución de las ecuaciones planteadas por el método de ML, cuando se aplica métodos numéricos (Maity 2018).

2.11 MODELOS UNIVARIADAS

Distribución Gamma; La variable aleatoria, no negativa, gamma, tiene muchas aplicaciones en recursos hídricos, debido a que numerosas variables presentan asimetría a la derecha, por lo que este modelo es utilizado en el análisis probabilístico de sequías (Tosunoğlu y Onof 2017). La función de distribución de probabilidad general de la variable aleatoria gamma se define de la siguiente forma:

$$gam(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1} \quad x \geq 0, \quad \alpha > 0, \quad \beta > 0$$

Donde: α es el parámetro de forma y $\frac{1}{\beta}$ es el de escala.

La función de distribución de probabilidad Gamma toma una amplia variedad de formas, las que depende del valor de sus parámetros, siendo la función exponencial y chi cuadrado las más importantes. Los momentos estadísticos de la variable aleatoria $gam(x; \alpha, \beta)$ de orden uno y dos son respectivamente:

$$E[x] = \frac{\alpha}{\beta} \qquad Var[x] = \frac{\alpha}{\beta^2}$$

Distribución Log-normal; Karamouz et al. (2012) mencionan que la distribución log-normal es una distribución continua, comúnmente utilizada en el análisis de eventos hidrológicos, cuando las variables aleatorias no pueden ser negativas. La distribución Log-normal es la distribución de una variable aleatoria cuyo logaritmo tiene el comportamiento normal. Así mismo, Maity (2018) menciona que el producto de muchas variables aleatorias independientes y positivas, puede resultar en una distribución Log-normal. Esto se justifica considerando el teorema del límite central, en el dominio logarítmico. La función de distribución de probabilidad de la variable aleatoria Log-normal es:

$$\text{logn}(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(x) - \mu)^2}{4\sigma^2}\right) \text{ para } x > 0$$

Esta función se puede derivar a partir de la función de distribución de probabilidad normal, la escala original se puede calcular a partir de la variable transformada logarítmicamente. Para calcular los momentos estadísticos de la variable x , se realiza sobre los logaritmos de x , se utilizan las siguientes ecuaciones:

$$E(x) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$\text{Var}(x) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

Distribución Gumbel; Maity (2018) menciona que la distribución de Gumbel a menudo se usa para modelar variables temporales, debido a que sus parámetros en la distribución proporcionan una gran flexibilidad para modelar sistemas en los que el número de fallas aumenta con el tiempo, disminuye con el tiempo o permanece constante con el tiempo. La distribución de Gumbel se desarrolló originalmente para la estimación de inundaciones; sin embargo, se adoptó en estudios de análisis de sequía (Dalezios et al. 2000). La función de distribución probabilística de Gumbel es:

$$f(x) = \frac{\beta}{\delta} \left(\frac{x}{\delta}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\delta}\right)^\beta\right), \text{ para } x > 0$$

Los momentos estadísticos de la variable aleatoria Gumbel son respectivamente:

$$E(x) = \delta \Gamma\left(1 + \frac{1}{\beta}\right)$$

$$Var(x) = \delta^2 \Gamma\left(1 + \frac{2}{\beta}\right) - \delta^2 \left(\Gamma\left(1 + \frac{1}{\beta}\right)\right)^2$$

Distribución Pearson tipo III; La distribución Pearson Tipo III es una distribución continua que se utiliza para representar observaciones sesgadas; por lo que, se usa ampliamente en estudios hidrológicos, como en inundaciones (Maity 2018).

$$f(x) = \frac{(\ln x - x_0)^{\gamma-1} \exp\left(-\frac{\ln x - x_0}{\beta}\right)}{x\beta^\gamma \Gamma(\gamma)}$$

Donde x_0 representa el parámetro de posición; β el parámetro de escala y γ es el parámetro de forma. Los momentos estadísticos de la variable aleatoria Pearson tipo III son respectivamente:

$$E(x) = x_0 + \gamma\beta \quad Var(x) = \gamma\beta^2$$

Distribución Weibull; La distribución Weibull se usa a menudo en la teoría de confiabilidad para modelar las fallas en un sistema, cuando la razón de fallos es proporcional a la potencia del tiempo, la función de distribución de probabilidad de la variable aleatoria Weibull es el siguiente:

$$f(x) = \frac{\beta}{\lambda} \left(\frac{x}{\lambda}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\lambda}\right)^\beta\right), \text{ para } x > 0$$

Los momentos estadísticos de la variable aleatoria Pareto x son respectivamente:

$$E[x] = a^{\frac{1}{b}} \Gamma(1+b^{-1}) \quad Var[x] = a^{\frac{-2}{b}} \left(\Gamma(1+2b^{-1}) - (\Gamma(1+b))\right)^2$$

Distribución exponencial; La distribución exponencial se ha venido usando para describir los tiempos de intensidad y severidad de las sequías (Chow et al. 1994). Otros estudios también usan esta distribución en el análisis de sequías modelando la duración y severidad, tal como reportan Shiau et al. (2007), Shiau y Shen (2001), Shiau y Modarres (2009). Esta distribución se adapta bien a variables discretas y continuas, por lo que es útil en el análisis bivariado de sequías (Chen y Guo 2019).

La función densidad de probabilidad de la distribución exponencial para la variable aleatorio x está dada por:

$$\exp(x, \lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x > 0; \lambda > 0 \\ 0, & \text{de otra manera} \end{cases}$$

Donde: λ es igual a $\frac{1}{\beta}$, mientras su función de distribución de probabilidad acumulada está dada por:

$$P(0 \leq X \leq x) = 1 - e^{-\lambda x}$$

$$F(x) = 1 - e^{-\lambda x}$$

La función densidad y la distribución acumulada se escribe de las siguientes formas:

$$f(x) = \frac{1}{\alpha} \exp\left(\frac{x-\beta}{\alpha}\right) \quad F(x) = 1 - \exp\left(\frac{x-\beta}{\alpha}\right)$$

Donde: α representa el parámetro de forma y β es el parámetro de ubicación o localización de la distribución exponencial. Estos parámetros pueden ser estimados por el método de los momentos MOM o estimadas con ML (Chow et al. 1988). Los momentos estadísticos de la variable aleatoria x de orden uno y dos son respectivamente:

$$E(x) = \frac{1}{\lambda} \quad Var(x) = \frac{1}{\lambda^2}$$

Distribución Pareto; La distribución de Pareto a menudo se describe como la base de la regla 80/20, también conocido como el principio de Pareto. La que establece que, para muchos eventos, aproximadamente el 80 % de los efectos provienen del 20 % de las causas, la distribución de Pareto se puede ver como la versión continua de la variable aleatoria discreta Zeta, mientras que la función de distribución de probabilidad de la variable aleatoria de Pareto, que es ampliamente conocida por su cola muy larga, se define de la siguiente manera (Grami 2017):

$$pareto(x; \alpha, \kappa) = \frac{\alpha \kappa^\alpha}{x^{\alpha+1}} \quad \alpha > 0 \text{ y } x \geq \kappa > 0$$

Los momentos estadísticos de la variable aleatoria Pareto x son respectivamente:

$$E(x) = \frac{\alpha k}{\alpha - 1} \quad \alpha > 1 \quad \text{Var}(x) = \frac{\alpha k^2}{(\alpha - 1)^2 (\alpha - 2)} \quad \alpha > 2$$

2.11.1 Análisis multivariado de sequías

Shiau y Shen (2001) reportan que, debido a la naturaleza compleja de las sequías, el análisis univariado no proporcionar una evaluación exhaustiva. En consecuencia, el análisis de forma individual de la distribución de la duración y la severidad de las sequías no revela la correlación significativa entre ellas. Por lo tanto, el enfoque adecuado para describir las relaciones entre las características de la sequía, es mediante el análisis bivariado, a pesar de los complejos cálculos que se requiere. Sin embargo Sklar (1959) propuso las cópulas para facilitar el cálculo en el análisis bivariado, así mismo Chen y Guo (2019) propusieron diferentes métodos para investigar la distribución conjunta de las variables de la sequía como la duración y la severidad, duración e intensidad. Este método fue utilizada por Shiau, (2006), Shiau y Modarres (2009), Mirabbasi et al. (2012), Yoo et al. (2013) y Tosunoğlu y Onof (2017), en el análisis bivariado de las sequías.

2.11.2 Cópulas

Las cópulas son funciones bivariadas introducidas por Sklar (1959) para vincular distribuciones marginales univariadas con distribuciones multidimensionales. En el caso bivariado, el teorema de Sklar establece que, si dos variables aleatorias x e y siguen las funciones arbitrarias de distribución marginal $F_X(x)$ y $F_Y(y)$ respectivamente. Entonces, existe un arreglo C para combinar estas dos funciones de distribución marginal, en una sola estructura llamada copula. Este modelo en su forma general se representa de la siguiente manera:

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$$

2.11.3 Cópulas arquimedianas

Un tipo de cópula son las Arquimedianas, las que son ampliamente utilizadas en hidrología, por ejemplo, en el análisis de las características bivariadas de las sequías. Las cópulas arquimedianas más utilizadas debido a su simplicidad y amplia representación son tres; Clayton, Gumbel y Frank. En la Tabla 3 se muestra las funciones de las cópulas y sus parámetros. La elección de la cópula que mejor se ajusta

a la relación entre dos variables depende del rango de niveles de dependencia que pueden describir dichas variables. Por ejemplo, Gumbel solo se puede aplicar para la dependencia positiva, mientras que Clayton y Frank son adecuados para las dependencias positivas y negativas (Chen y Guo 2019).

La relación entre el parámetro de las funciones arquimedianas (θ) y la tau de Kendall para las tres cópulas se muestran en la Tabla 3. θ es el parámetro utilizado para medir el grado de asociación entre las variables univariadas u y v . El coeficiente de correlación de Kendall, comúnmente conocido como tau de Kendall, se usa para medir la dependencia entre dos variables aleatorias.

Tabla 3: Cópulas arquimedianas

Tipo de copula	Función	Relación entre τ y θ
Clayton	$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$\tau = \frac{\theta}{\theta + 2}, \theta \in [1, \infty)$
Gumbel	$C(u, v) = \exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right\}$	$\tau = 1 - \frac{1}{\theta}, \theta \in (0, \infty)$
Frank	$C(u, v) = -\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right]$	$\tau = 1 + \frac{4}{\theta} \left[\frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} d(t) - 1\right], \theta \in R$

Nota; τ es la tau de Kendall, θ parámetros de las funciones arquimedianas, u, v son las funciones de distribución acumuladas de las variables correlacionadas.

2.12 PERIODO DE RETORNO Y RIESGO

De acuerdo con Maity (2018) el período de retorno también conocido como intervalo de ocurrencia promedio se determina para cualquier evento hidrológico como inundaciones, precipitaciones, descargas de ríos, deslizamientos de tierra, tormentas de viento, tornados, etc. Este periodo de retorno es importante en el análisis de riesgo y la incertidumbre en estudios hidroclimáticos. Se define formalmente como el tiempo promedio para que un evento de una magnitud dada sea igualado o excedido en un sentido estadístico. Básicamente es la medida estadística calculada sobre datos históricos.

2.12.1 Evento

Un evento es un suceso que se describe como una ocurrencia de interés y acontece cuando la variable aleatoria X , característica de un proceso hidrológico, es igual o mayor a un umbral arbitrario, también llamado nivel de excedencia T_x y se define como:

$$X \geq x_T$$

La magnitud de un evento raro o poco probable corresponde a un umbral muy alto, el cual es inversamente proporcional a su probabilidad de ocurrencia. Los eventos de gran magnitud ocurren con menos frecuencia que los eventos moderados o leves. Además, el análisis de frecuencia tiene como objetivo vincular un evento hidrológico con su frecuencia de ocurrencia por medio de distribuciones de probabilidad.

2.12.2 Intervalo de ocurrencia

El intervalo de ocurrencia, es el periodo que separa dos eventos sucesivos, se ha denotado por τ .

2.12.3 Periodo de retorno

El periodo de retorno T de un evento ($X \geq x_T$), es el promedio de la variable aleatoria τ , que representa la medida del intervalo de ocurrencia, obteniéndose para un gran número de ocurrencias el valor esperado de τ :

$$T(x_T) = E[\tau]$$

En términos de probabilidad, solo dos casos son posibles para el evento de interés

- a) El evento ocurre, si $X \geq x_T$ y se le denomina éxito, con probabilidad de p .
- b) El evento no ocurre, si $X < x_T$ y se le denomina falla, con probabilidad de $1-p$.

Entonces, el resultado de un ensayo solo puede ser, éxito o falla. Por lo que se trata de una variable aleatoria dicotómica, correspondiendo a un ensayo Bernoulli. Por esta razón, el cálculo de la probabilidad de ocurrencia para un número de eventos, estará definida por la multiplicación de $(y-1)$ consecutivas fallas seguido por el evento éxito.

Como los ensayos son independientes, se trata del proceso Bernoulli, siendo su probabilidad:

$$\underbrace{(1-p)(1-p)\dots(1-p)}_{y-1} p = (1-p)^{y-1} p$$

Lo que vendría a ser la función de frecuencia aleatoria Geométrica con parámetro p , cuya variable aleatoria se define como el número de ensayos necesarios para que ocurra el suceso favorable. Entonces, para determinar el periodo de retorno, se calcula el valor esperado de la función Geométrica, de acuerdo con la siguiente expresión:

$$E[\tau] = \sum_{y=1}^{\infty} y(1-p)^{y-1} p$$

$$E[\tau] = \frac{1}{p}$$

El periodo de retorno de un evento ($X \geq x_T$) se puede relacionar con la probabilidad de ocurrencia de un evento, asumiendo que la ocurrencia de probabilidad del evento $X \geq x_T$ en cualquier año es $P(X \geq x_T)$ y considerando la ecuación $E[\tau]$, de la siguiente forma:

$$T(x_T) = \frac{1}{P(X \geq x_T)} = \frac{1}{1 - P(X < x_T)} = \frac{1}{p(x_T)} = \frac{1}{1 - F(x_T)}$$

Donde: $T(x_T)$ representa el periodo de retorno y está definida como la inversa de la probabilidad de excedencia $P(x_T) = 1 - F(x_T)$, donde $F(x_T)$ es la función acumulada de la distribución con mejor ajuste. Por lo que cuanto más raro es el evento, menor es su probabilidad de superación. Por lo tanto, mayor es su período de retorno, si se considera un período de tiempo muy largo, el evento con un valor superior a x_T ocurrirá en promedio una vez cada T años.

El período de retorno, es el valor esperado del intervalo de ocurrencia, posee una unidad de medida dado en año, trimestre, mes o cualquier otra medida de duración. En hidrología, el período de referencia suele ser un año, correspondiente a un ciclo hidrológico. Cuando se trabaja con series interanuales, para aumentar la cantidad de información, la distribución debe corregirse para que vuelva al período anual, mediante la siguiente relación.

$$T(x_r) = \frac{N}{n[1 - F(x_r)]}$$

Donde: N es el total de la longitud de la serie en años, n número de eventos en el total de la serie.

2.13 PERIODO DE RETORNO DE EVENTOS DE SEQUÍA

El enfoque para un buen diseño hidráulico e hidrológico es el análisis del período de retorno de los eventos hidrológicos, este análisis es usada para calcular la probabilidad de ocurrencia de un evento determinado en un periodo determinado (Chow et al. 1988). En el caso de las sequías, el período de retorno proporciona información útil para la gestión del uso adecuado del agua en periodos de escasez, El período de retorno de una sequía puede definirse como el tiempo promedio transcurrido o el intervalo medio entre las ocurrencias de dos eventos (Shiau y Shen 2001).

2.13.1 Período de retorno univariado de sequías.

El período de retorno para la serie de duración parcial con distribución univariado, se puede derivar basado en el concepto de proceso estocástico bajo el supuesto de que los eventos están distribuidos de forma independiente e idéntica. Por lo que el período de retorno de un evento de sequía con una duración d , con una severidad s o con una intensidad i , se puede calcular mediante las relaciones dadas por Shiau y Shen (2001):

$$T(d) = \frac{E(Tes)}{1 - F(D)} = \frac{N(E(Tes))}{n[1 - F(D)]}$$

$$T(s) = \frac{E(Tes)}{1 - F(S)} = \frac{N(E(Tes))}{n[1 - F(S)]}$$

$$T(i) = \frac{E(Tes)}{1 - F(I)} = \frac{N(E(Tes))}{n[1 - F(I)]}$$

Donde: $E(Tes)$ representa el valor esperado del tiempo entre sequías, contado desde el inicio de la sequía hasta el inicio de la siguiente sequía. $T(d)$ es el periodo de retorno de la duración de sequías, $T(s)$ es el período de retorno para la severidad de la sequía y $T(i)$ es el período de retorno para la intensidad de la sequía, $F(D)$, $F(S)$ y $F(I)$

son las funciones de distribución acumuladas de las variables aleatorias, duración, severidad e intensidad respectivamente. N es longitud total de la serie y n número de eventos de sequía.

2.13.2 Periodo de retorno bivariado de sequías.

Shiau (2006) definió el período de retorno conjunto o bivariado, para la relación entre la duración y la severidad de un evento de sequía, mediante la siguiente ecuación:

$$T_{Dys} = \frac{E(Tes)}{1 - P(D \geq d \wedge S \geq s)} = \frac{E(Tes)}{1 - F(D) - F(S) + C(F(D), F(S))}$$

Donde: T_{Dys} simboliza el periodo de retorno bivariado conjuntivo para un evento de sequía con una duración D mayor igual a una duración d y una severidad S mayor igual a s , definida por la condición $(D \geq d \wedge S \geq s)$, de igual modo definió el periodo de retorno bivariado disyuntivo, de la siguiente forma:

$$T_{DoS} = \frac{E(Tes)}{1 - P(D \geq d \vee S \geq s)} = \frac{E(Tes)}{1 - C(F(D), F(S))}$$

Donde: T_{DoS} representa el periodo de retorno bivariado disyuntivo para un evento de sequía con una duración D mayor igual a d o una severidad S mayor igual a s , definida por la condición $(D \geq d \vee S \geq s)$. Además, $C(F(D), F(S))$ es la cópula que mejor ajuste tiene las funciones $F(D)$ y $F(S)$, que puede ser Clayton, Gumbel y Frank. En esta misma línea de investigación, Chen y Guo (2019) extendieron esta fórmula para el periodo de retorno bivariado para la duración e intensidad:

$$T_{DyI} = \frac{E(Tes)}{1 - P(D \geq d \wedge I \geq i)} = \frac{E(Tes)}{1 - F(D) - F(I) + C(F(D), F(I))}$$

Donde: T_{DyI} representa el periodo de retorno para un evento de sequía con una duración D mayor e igual a d y una intensidad I mayor e igual a i $(D \geq d \wedge I \geq i)$. También formularon el periodo de retorno bivariada disyuntivo para la duración e intensidad, de la siguiente manera:

$$T_{DoI} = \frac{E(Tes)}{1 - P(D \geq d \vee I \geq i)} = \frac{E(Tes)}{1 - C(F(D), F(I))}$$

Donde: $T_{D \vee S}$ representa el periodo de retorno para un evento de sequía con una duración D mayor e igual a d o una intensidad I mayor e igual a i ($D \geq d \vee S \geq s$). Además $c(F(D), F(I))$ es la cópula Clayton, Gumbel y Frank, que mejor ajuste tiene.

2.14 PRUEBAS ESTADÍSTICAS

2.14.1 Prueba de Ljung-Box.

Para validar la efectividad de las predicciones del modelo ARMA se debe analizar los residuos estimados, los cuales se deben comportar como un ruido blanco; con distribución normal, media cero, covarianza nula y constante. la prueba estadística Ljung-Box permite evaluar la independencia de los m primeros residuos evaluados, el estadístico de Ljung-Box se plantea de la siguiente forma (Salas et al. 1980).

$$Q(m) = T(T+2) \sum_{k=1}^m \frac{\hat{p}_k^2}{n-r}$$

Donde:

\hat{p}_k = coeficiente de autocorrelación de los residuos estimados
 T = número de valores de la serie
 r = número de parámetros estimados
 m = rezagos en la autocorrelación

Este estadístico tiene como Hipótesis nula que, $Q(m)$ se distribuye como un modelo Chi-cuadrado con un número de grados de libertad igual al número de rezagos considerados, menos el número de parámetros estimados, menos 1 ($m-r-1$).

2.14.2 El error cuadrático medio (RMSE)

El valor RMSE mide las diferencias entre los valores observados y estimados. El RMSE se puede expresar como la siguiente ecuación:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - x)^2}$$

Donde x_i es los valores estimados y x representa los valores observados, El RMSE otorga un peso relativamente mayor a los errores grandes al cuadrarlos.

III. MATERIALES Y MÉTODOS

3.1 DESCRIPCIÓN DEL ÁREA DE ESTUDIO

La cuenca del río Mantaro está ubicada en la parte central y occidental de América del Sur, en los Andes centrales del Perú, pertenece a la región húmeda y drena una hoya hidrográfica de 34 400 km². El río Mantaro se forma en la laguna de Junín, ubicada a una altura de 4083 m.s.n.m., recorre una distancia de 724 km de norte a sur predominantemente, llegando a descender a la altura de 476 m.s.n.m., en la localidad de Boca Mantaro, donde finalmente confluye con el río Apurímac dando lugar al río Ene (Figura 4).

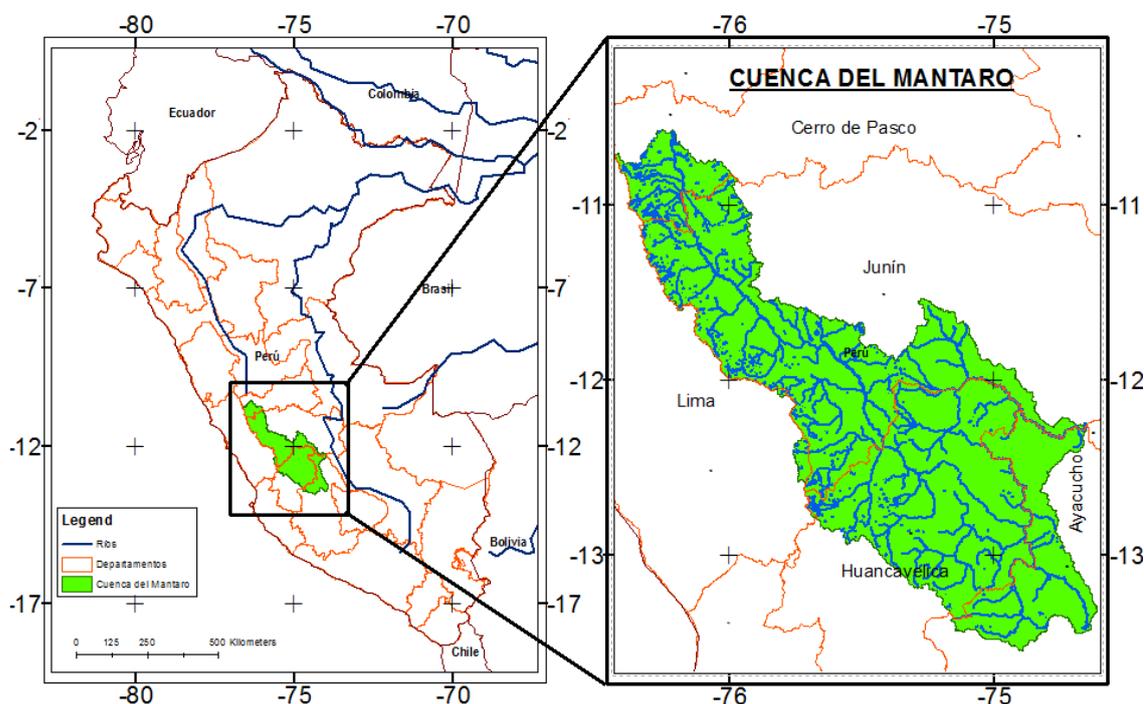


Figura 4: Mapa de ubicación de la cuenca hidrográfica del río Mantaro.

Fuente Elaboración Propia.

3.1.1 Ubicación Geográfica

Geográficamente se ubica entre las coordenadas 10° 30' a 13° 30' de Latitud Sur y entre las 74° 00' a 76° 30' de Longitud Oeste. Abarca los departamentos de Pasco,

Junín, Huancavelica y Ayacucho, ocupando las provincias de Pasco, Junín, Yauli, Jauja, Concepción, Huancayo, Tayacaja, Huancavelica, Acobamba, Angaraes, Huanta y Huamanga. Esta cuenca forma parte de la cuenca del río Amazonas y de la vertiente del océano Atlántico, sus límites son por el norte, la cuenca Alto Huallaga; por el oeste, las cuencas Huaura, Huaral, Chillón, Rímac, Mala, Cañete, San Juan y Pisco; al sur, las cuencas de Pampas y Bajo Apurímac; al este, con las cuencas del Perene y Alto Amazonas.

3.1.2 Climatología

El clima en la cuenca del Mantaro es variado. El rango anual de la temperatura media oscila entre los 2,1 °C para toda la cuenca, los meses de enero y febrero presentan las máximas temperaturas, mientras que, los meses de junio y julio presentan las temperaturas más bajas durante el año. Las precipitaciones en la cuenca del Mantaro se inician en el mes de agosto y se incrementan gradualmente en los meses de setiembre y octubre, haciéndose más recurrentes a partir de noviembre, alcanza valores máximos los meses de enero, febrero y marzo, para luego disminuir bruscamente en el mes de abril, alcanzando su valor mínimo en el mes de junio, De acuerdo con el informe de MINAG (2010) la cuenca del Mantaro cuenta con 8 zonas de vida de las 38 zonas de vida consideradas por Holdridge, además, estas zonas se subdividen en 38 pisos ecológicos.

3.2 MATERIALES

En este estudio, se consideró los registros de precipitación mensual de 43 estaciones meteorológicas ubicadas en la cuenca del Mantaro. En la Tabla 4 se muestra las estaciones seleccionadas, indicando su ubicación, la altura sobre el nivel del mar a la que se encuentra, año de inicio y fin del registro, número de datos faltantes y número de datos del registro.

Tabla 4: Estaciones pluviométricas

Nº	Nombres	Lat.	Lon.	Alt.	Años ini	Años fin	DF	DT
1	Acobamba	-12.864	-74.559	3337	1965	2011	57	495
2	Atocsaico	-11.299	-76.078	4200	1952	2001	12	576
3	C.h. malpaso	-11.410	-76.042	3795	1940	2001	2	730
4	Carhuamayo	-10.924	-76.057	4130	1952	2001	0	588
5	Casapato	-10.981	-76.039	4100	1952	2001	11	577
6	Cercapuquio	-12.380	-75.303	4390	1965	2005	58	422
7	Chilicocha	-12.701	-75.452	4546	1965	2005	32	448
8	Choclococha	-13.100	-75.074	4696	1965	2001	21	411
9	Corpacancha	-11.365	-76.215	4150	1956	2001	42	498
10	Huamanga	-13.150	-74.217	2773	1965	2008	128	388
11	Huancalpi	-12.542	-75.237	4438	1965	2010	21	519
12	huancavelica	-12.779	-75.034	3689	1965	2008	154	362
13	Huayao	-12.034	-75.339	3357	1972	2016	8	520
14	Hueghue	-11.237	-76.281	4172	1965	2016	12	600
15	Huichicocha	-12.562	-75.529	4693	1965	2005	19	461
16	Ingenio	-11.881	-75.288	3200	1965	2010	71	469
17	Jauja	-11.779	-75.474	3361	1970	2009	84	384
18	La oroya	-11.576	-75.966	4007	1995	2013	0	216
19	La quinua	-13.034	-74.135	3316	1965	2010	79	461
20	Lircay	-12.983	-74.729	3280	1965	2010	9	531
21	Luricocha	-12.817	-74.234	2589	1965	1999	50	358
22	Marcapomacocha	-11.405	-76.325	4437	1969	2013	40	488
23	Matibamba	-12.084	-74.817	2191	1965	1984	21	207
24	Morococha	-11.601	-76.135	4505	1933	2001	53	763
25	Pachacayo	-11.805	-75.719	3616	1966	2005	6	462
26	Pachachaca	-11.617	-76.034	4081	1997	2013	0	192
27	Palaco	-12.347	-75.299	3542	1965	2001	17	415
28	Pampas	-12.393	-74.866	3232	1989	2011	15	249
29	Paragsha	-10.674	-76.257	4310	1952	2001	32	556
30	Paucarbamba	-12.467	-74.567	3370	1965	2011	164	388
31	Pilchaca	-12.406	-75.085	3650	1965	2011	59	493
32	Presa upamayo	-10.925	-76.278	4105	1952	2001	0	588
33	Salcabamba	-12.199	-74.786	3275	1965	2011	117	435
34	San pedro de Cachi	-13.012	-74.399	2990	1965	2010	25	515
35	Shelby	-10.810	-76.225	4120	1955	2001	8	544
36	Tambo del sol	-10.874	-76.119	4119	1969	2005	12	420
37	Telleria	-12.375	-75.121	3050	1965	2005	20	460
38	Upamayo	-10.925	-76.278	4111	1965	2005	0	480
39	Viques	-12.152	-75.237	3186	1965	2010	57	483
40	Wallyapampa	-13.068	-74.220	2470	1966	2010	169	359
41	Yatac	-11.334	-76.400	4688	1969	2013	25	503
42	Yauli	-11.667	-76.084	4120	1997	2013	0	192
43	Yauricocha	-11.956	-75.907	4421	1965	2005	2	478

Nota: Lat; Latitud, Log; longitud, Alt; altura sobre el nivel del mar, Años ini; años de inicio, Años fin; años fin, DF; número de meses faltantes y DT; número total de datos.

3.2.1 Programas de cómputo utilizados.

R (Lenguaje de programación); R es un lenguaje de programación para el análisis estadístico, este programa es la versión mejorada del programa libre S y es libre. En la actualidad R es uno de los lenguajes de programación más utilizados en investigación científica, siendo además muy popular en los campos de aprendizaje automático, minería de datos, investigación biomédica, bioinformática y matemáticas financieras. A esto, se añade la posibilidad de cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo. En la presente tesis se ha utilizado el paquete SPEI; paquete que tiene conjunto de funciones para calcular la evapotranspiración potencial y varios índices de sequía, incluido el índice estandarizado de precipitación y evapotranspiración (Beguería y Vicente-Serrano 2013).

Matlab; llamado así por su abreviatura de MATrix LABoratory, es un sistema de cómputo numérico que ofrece un entorno de desarrollo integrado con un lenguaje de programación propio, llamado M. Está disponible para las plataformas Unix, Windows, Mac OS X y GNU/Linux (MATLAB 2020). En la presente tesis se ha utilizado para la generación de datos usando el modelo estocástico ARMA, así como también en el análisis de los eventos de sequías, para determinar la mejor distribución univariada y bivariada, así como también el cálculo del periodo de retorno bivariado.

Qgis; anteriormente llamado Quantum GIS, es un programa libre para sistema de información geográfica, para plataformas GNU/Linux, Unix, Mac OS, Microsoft Windows y Android. Permite manejar formatos ráster y vectoriales a través de la biblioteca GDAL, así como también, bases de datos (Yolanda y Alfonso 2015). En la presente tesis se ha utilizado para delimitar la cuenca del Mantaro, realizar el análisis de polígono de Thiessen y la elaboración de mapas.

3.3 METODOLOGÍA

La presente tesis sigue una secuencia de cinco etapas o fases, las que se muestran en el diagrama de flujo presentado en la Figura 5, cada una de estas etapas se detalla a continuación.

Primera etapa; en esta etapa se realiza la inspección visual de los datos mediante el gráfico de las series, con la finalidad de detectar datos atípicos, tendencias y datos faltantes. También se analiza la homogeneidad y tendencias de las series aplicando

cuatro pruebas; Pettitt, ratio Von-Neuman, Bushand y SNHT. Las que fueron descritas en la sección 3.7.2. Se aplica las cuatro pruebas por ser complementarias y permiten determinar la homogeneidad en la serie de tiempo, así como también, la identificación de saltos que pudiera presentarse, también determina el punto de ruptura de la serie. Además, se aplicó la prueba de Mann-Kendall, para determinar posibles tendencias en las series de precipitación mensual. Finalmente, en esta etapa se identifica el número de datos de cada serie. Teniendo en cuenta que para aplicar el modelo autorregresivo se requiere al menos 50 observaciones como número mínimo de datos, así lo indica Pankratz (2009).

Segunda etapa; en esta etapa se determina las zonas homogéneas de la cuenca del Mantaro, aplicando el método de clúster, para lo cual se considera como criterio de agrupamiento la latitud y longitud de cada estación meteorológica, así como también su altura sobre el nivel del mar y el promedio mensual de precipitación. Esta primera agrupación se mejora, aplicando el método del polígono de Thiessen, dando como resultado la identificación de las zonas homogéneas de la cuenca. También, se calcula el promedio de las series ubicadas en cada zona identificada, para obtener una sola serie de precipitación mensual que representara a cada zona homogénea identificada en la cuenca.

Tercera etapa; en esta etapa se identifican y se usan los modelos autorregresivos para generar 500 años de series de precipitación sintéticas, basados en las series promedio en cada zona homogénea. Considerando el primer paso en esta etapa, la transformación o estandarización de la serie de precipitación mensual, con el objetivo de obtener una serie con media cero y varianza dada. Paso seguido, se determina el orden del modelo ARMA(p,q) analizando las funciones de autocorrelación (FAC) y las funciones de autocorrelación parcial (FACP), la identificación del orden del modelo es mejorada aplicando el criterio de BIC. Paso seguido se estima los parámetros de los modelos y se realiza la transformación inversa de la serie. En esta etapa se realiza dos pruebas de validación: la primera, es la validación del modelo mediante la contrastación de serie original y la observada, para ello se utiliza la prueba RMSE, coeficiente de determinación y de regresión; La segunda, es la prueba de validación de la serie pronosticada mediante la prueba estadística Ljung-Box y Q-Test.

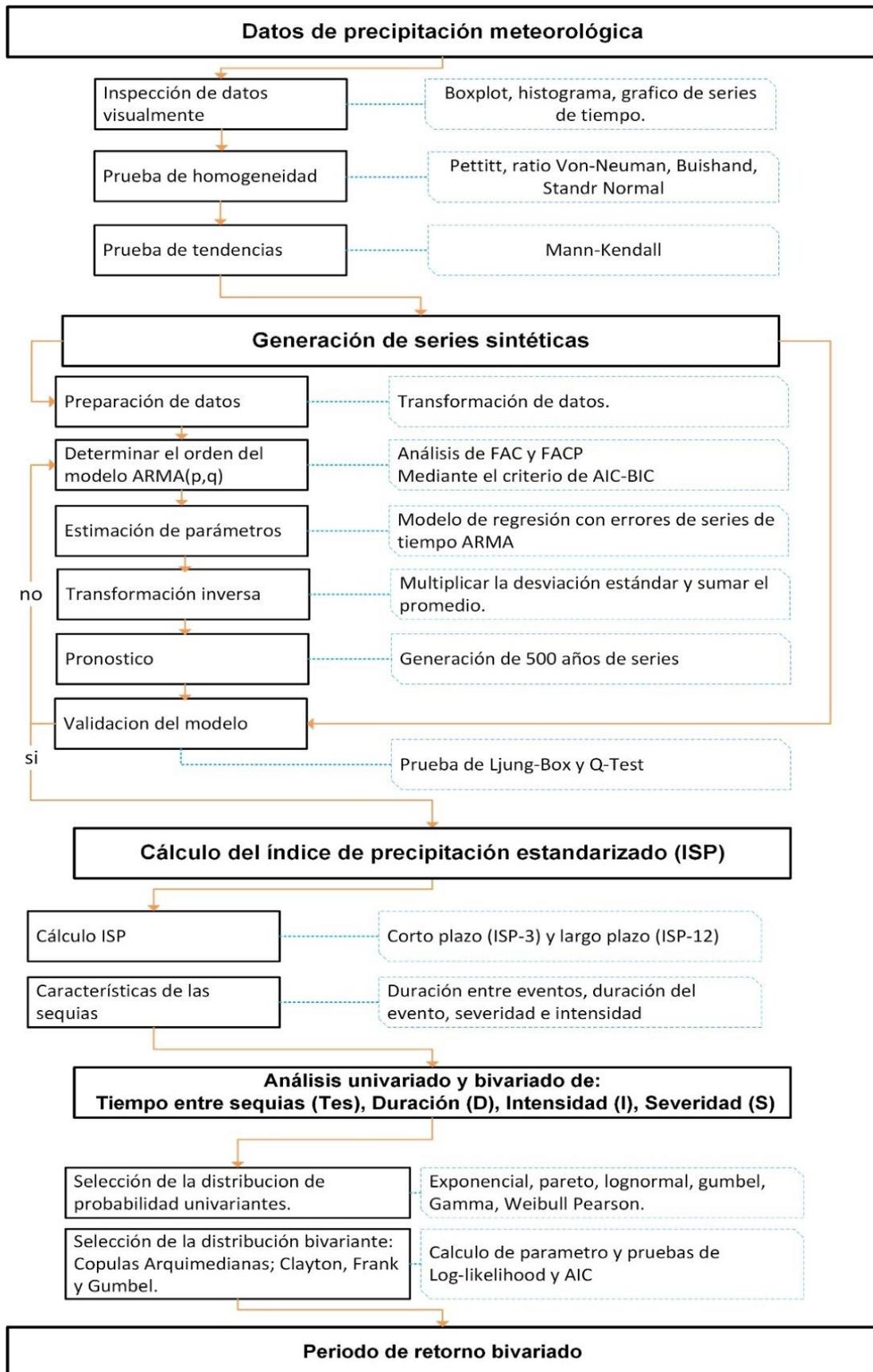


Figura 5: Diagrama de flujo de la metodología, Fuente: Elaboración propia

Cuarta etapa; en esta etapa se identifica el inicio, fin, severidad e intensidad de las sequías identificadas en cada zona homogénea. Basados el análisis de los resultados del índice de precipitación estandarizado, llamada serie SPI. Además, se identifica las sequías utilizando los niveles de truncamiento expuestos en la Tabla 1, para lo cual se escribió un script en Matlab teniendo en consideración las definiciones sobre caracterización de sequías dadas en la sección 3.6.

Quinta etapa; en esta etapa se realiza el análisis univariado y bivariado de los eventos de sequías identificadas en la etapa anterior. En el análisis univariado se consideran como variables aleatorias al tiempo entre sequías (Tes), duración de la sequía (D), intensidad de las sequías (I) y severidad de las sequías (S). Estas variables son ajustadas a las distribuciones teóricas como Gamma, Log normal, Gumbel, Pearson, Weibull, Exponencial y Pareto, escogiendo la distribución con mejor ajuste para cada variable. En el análisis bivariado se consideran tres cópulas arquimedianas; Clayton, Frank y Gumbel. Para determinar la mejor copula, primero se determina los parámetros y se prueba la calidad de ajuste, usando la prueba estadística de Log-likelihood y AIC.

Sexta etapa; finalmente se ha calculado el periodo de retorno de las variables duración y severidad, tanto de manera univariada y bivariada. Para las sequías a largo plazo (SPI-3) y a largo plazo (SPI-12).

IV. RESULTADOS Y DISCUSIÓN

4.1 ANÁLISIS DE LA HOMOGENEIDAD Y TENDENCIA

El análisis de la homogeneidad se realizó aplicando cuatro pruebas: Von Neumann (V-N), Buishand (Bh), homogeneidad normal estándar (SNHT) y Pettitt (Pt), para evaluar la tendencia se aplicó la prueba de Mann Kendall (M-K). En el Anexo 8.1, se muestran los resultados del análisis de homogeneidad y tendencia aplicada a cada serie. Obteniéndose los p-valor cercanos al cero para la estación Choclococha, para las pruebas de Bh, H-N y Pt, indicando que esta serie no es homogénea, de igual forma, el p-valor para la prueba de M-K es de 0.02, rechazándose la Hipótesis de no tendencia de la serie a un nivel de significancia de 1 %; por lo que, esta serie no se considera en análisis subsiguientes. Existen otros tres casos similares: Salcabamba, San Pedro de Cachi y Yatac, que también son excluidos del presente estudio, debido a la no homogeneidad y por presentar tendencias.

4.2 DETERMINACIÓN DE REGIONES HOMOGÉNEAS

Las características físicas y climáticas de la cuenca del Mantaro son variadas, dentro de su espacio diverso se encuentran dispersas las 46 estaciones meteorológicas en estudio, por lo que, para realizar el estudio adecuado, estas estaciones fueron agrupadas de acuerdo a la uniformidad climáticamente. Para lo cual, se aplicó primero el método clúster, basado en las variables de precipitación media anual, altura sobre el nivel del mar, longitud y latitud, obteniéndose una agrupación preliminar de las estaciones, el resultado de este análisis es el dendrograma que se muestra en la Figura 6, donde se muestra los cuatro conglomerados, agrupamiento que acontece al nivel de similitud de cuatro. El primer conglomerado; de izquierda a derecha, se compone de 23 estaciones meteorológicas ubicadas prevalentemente al norte de la cuenca, denominando este grupo como zona homogénea cuenca Alta; el segundo conglomerado, se compone de 15 estaciones meteorológicas ubicadas en la parte sur de la cuenca, nombrando esta zona homogénea como cuenca Baja; el tercer grupo se

compone de nueve estaciones meteorológicas, ubicada en la parte central oeste de la cuenca, denominando esta zona homogénea como cuenca media Oeste y finalmente el cuarto conglomerado se compone de 16 estaciones meteorológicas ubicadas en la parte central este de la cuenca, nombrándola zona homogénea, cuenca media Este, las dos últimas zonas corresponden a la cuenca media.

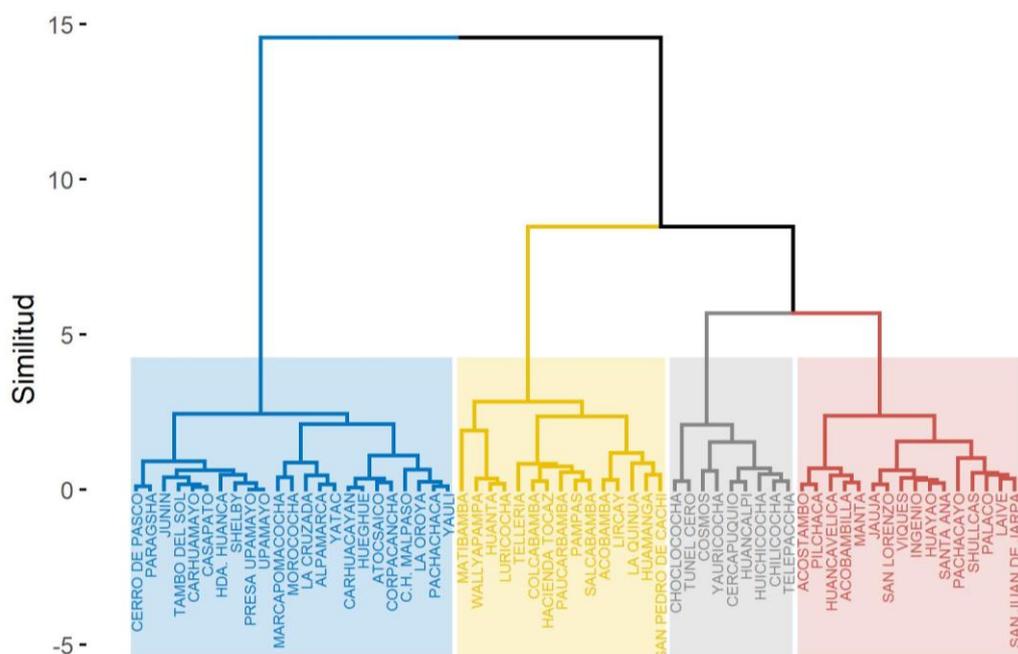


Figura 6: Dendrograma de regiones homogéneas, Fuente: elaborado en R.

La agrupación realizada concuerda con las condiciones físicas y climáticas de la cuenca del Mantaro, la que se observa en la Figura 7. Sin embargo, se identifican algunas inconsistencias, por ejemplo, la estaciones Meteorológica de Telleria ubicada en la cuenca media, se clasificó como parte del grupo de la cuenca Baja, debiendo ser ubicada en el grupo cuenca media Este. Otros casos similares, es la ubicación de las estaciones de Laive y Palaco, que de acuerdo con su ubicación deberían incluirse en la zona de la cuenca media Oeste, sin embargo, están clasificadas como parte de cuenca media Este. Lo mismo ocurre para la estación de Pachacayo, que por su ubicación debe pertenecer a la cuenca Alta, pero se encuentra ubicada en la cuenca Este.

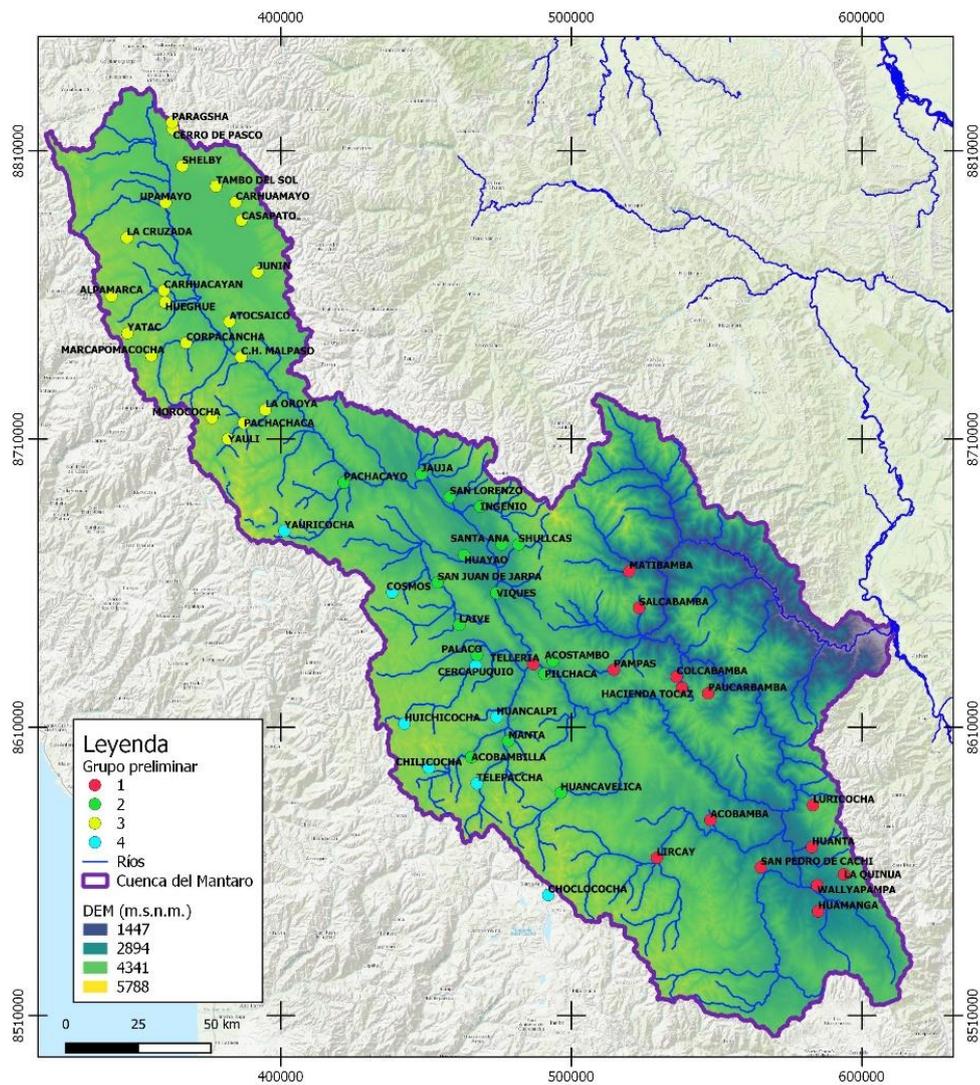


Figura 7: Distribución espacial preliminar de las estaciones agrupadas.

Para mejorar la agrupación de las estaciones meteorológicas, así como también, la identificación y delimitación de zonas homogéneas, se aplicó la metodología del polígono de Thiessen; integrado como rutina en el programa libre Qgis (Pokojski y Pokojka 2018). Este método consistió en delimitar áreas de influencia de cada estación ubicada en la cuenca, las que son consideradas unidades discretas. El tamaño y la configuración de los polígonos depende de la distribución de cada estación y para configurar las zonas homogéneas se juntan los polígonos de las estaciones pertenecientes al mismo grupo clasificado por el método clúster. Finalmente, se obtuvo cuatro zonas homogéneas visiblemente delimitadas; cuenca Alta, cuenca Baja, cuenca media Este y cuenca media Oeste, mostradas en la Figura 8.

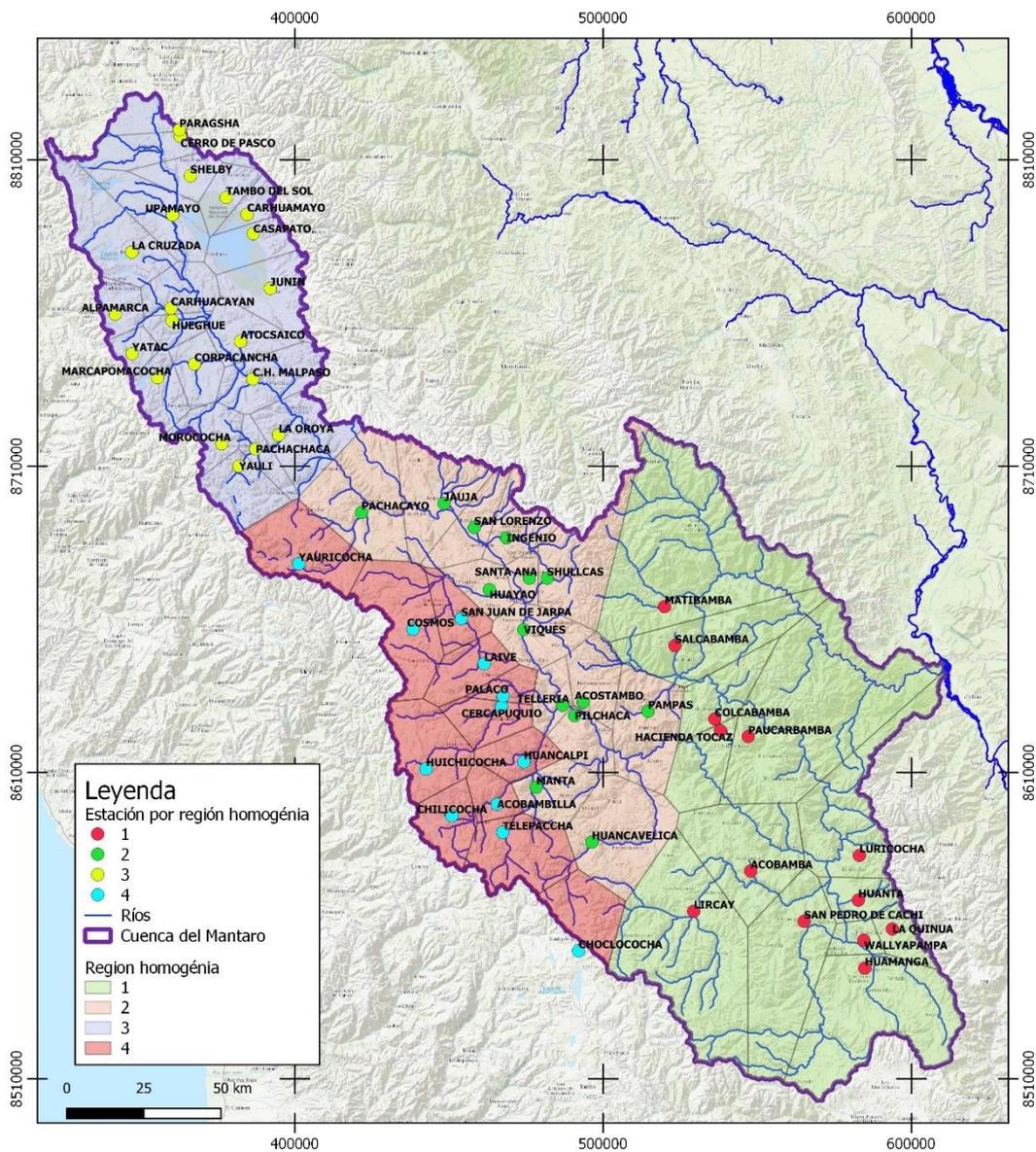


Figura 8: Delimitación de regiones homogéneas con el método de Thiessen

La zona homogénea cuenca Alta está constituida por las estaciones; Paragsha, Cerro de Pasco, Shelby, Tambo del Sol, Carhuamayo, Casapato, Upamayo, La Cruzada, Alpamarca, Carhaucayan, Hueghue, Atocsaico, Yatac, Marcapomacocha, Corpacancha, C.H. Malpaso, Morococha, La Oroya Pachachaca y Yauli. La zona homogénea cuenca media Oeste está constituida por las estaciones; Yauricocha, Cosmos, San Juan de Jarpa, Laive, Palaco, Cercapuquio, Huancalpi, Huichicocha, Chilicocha, Acobambilla y Telepaccha. La zona homogénea cuenca media Este está constituido por las estaciones; Pachacayo, Jauja, San Lorenzo, Ingenio, Santa Ana, Shullcas, Huayao, Viques, Acostambo, Pilchaca, Pampas, Manta y Huancavelica. La

zona homogénea cuenca Baja está constituida por las estaciones; Matibamba, Salcabamba, Colcabamba, Paucarbamba, Luricocha, Hacienda Tocas, Acobamba, Lircay, Huanta, San Pedro de Cachi, La Quinua, Wallyapampa, Huamanga.

4.3 HOMOGENEIZACIÓN DE DATOS POR ZONAS DE ESTUDIO

En cada zona homogénea identificada se halló el promedio de las series ubicadas dentro de cada zona, obteniéndose cuatro series representativas para cada zona identificada, las cuales se muestran en la Figura 9.

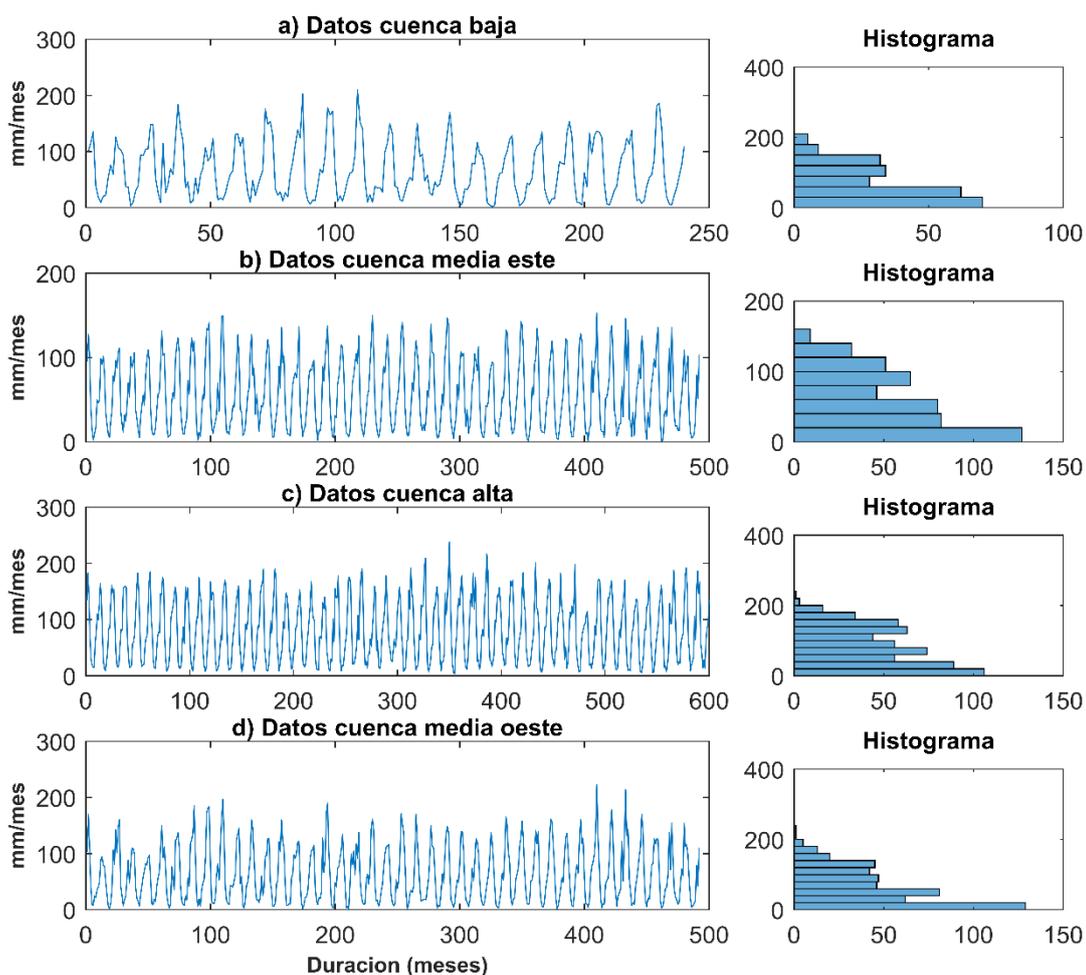


Figura 9: Series de tiempo precipitación.

Las cuatro series presentan una periodicidad debido a la estacionalidad climática propia de la cuenca del Mantaro. También se puede observar que las series son estacionarias: no presentan tendencias positivas ni negativas. Los histogramas de las series de precipitación en las cuatro zonas homogéneas obedecen aproximadamente a la distribución Gumbel. La serie de la cuenca Baja tiene un registro de precipitación mensual con 240 datos, mientras que la serie de la cuenca media Este cuenta con 492

datos, la serie de la cuenca Alta tiene un registro de 600 meses y finalmente la serie correspondiente a la cuenca media Oeste tiene un registro de 492 datos. Las cuatro series supera la cantidad de datos mínimos requeridos para ser modelados mediante modelos autorregresivos de medias móviles.

La evaluación visual de la homogeneidad y tendencia de las series es corroborada al aplicar las pruebas estadísticas V-N, Bh, SNHT y Pt. Los resultados de estas pruebas se muestran en la Tabla 5, los p -valores de las cuatro pruebas para las cuatro series son mayores al nivel de significancia del 0.05; por lo tanto, no se rechaza la Hipótesis nula, concluyendo que las series de precipitación son homogéneas, caso similar se encontró para los p -valores de la prueba M-K, concluyendo que las cuatro series no presentan tendencia.

Tabla 5: Prueba de homogeneidad y tendencia de las series

Región homogénea	Nº Datos	Prueba de homogeneidad (p-valor)				Prueba de tendencias M-K
		V-N	Bh	SNHT	Pt	
Cuenca Alta	600	0.494	0.959	0.4785	1.460	0.848
Cuenca media Este	492	0.541	1.000	0.509	0.755	0.337
Cuenca media Oeste	492	0.549	0.922	0.477	1.200	0.112
Cuenca Baja	240	0.633	0.763	0.795	0.755	0.105

Nota: V-N, prueba de Von Neumann; Bh, prueba de Buishand; SNHT, Standard Normal Homogeneity Test; Pt, prueba de Pettitt; M-K, prueba de Mann Kendall.

4.4 GENERACIÓN DE SERIES SINTÉTICAS

Analizar las sequías bajo el enfoque bivariado requiere de un número adecuado de datos en cada registro de precipitación, por tal motivo se generó una serie de precipitación de 500 años, basados en las series promedio de cada zona homogénea identificada, para lo cual se usó los modelos autorregresivos de medias móviles (Kim et al. 2006), siendo el procedimiento el siguiente.

4.4.1 Estandarización de las series

El primer paso fue estandarizar las cuatro series. Para tal efecto, se dividió con su varianza a la suma de las diferencias de la media con cada valor de la serie. El resultado es una serie temporal estacionaria, sin tendencia y con distribución normal: con media cero y varianza igual a σ^2 . El resultado de este proceso se presenta en la Figura 10.

Como se observa las cuatro series no presentan tendencias, no presentan estacionalidad y su histograma obedece a una distribución normal con media cero, por lo que reúnen las condiciones iniciales para ser modeladas usando los modelos autorregresivos de medias móviles (Salas et al. 1980).

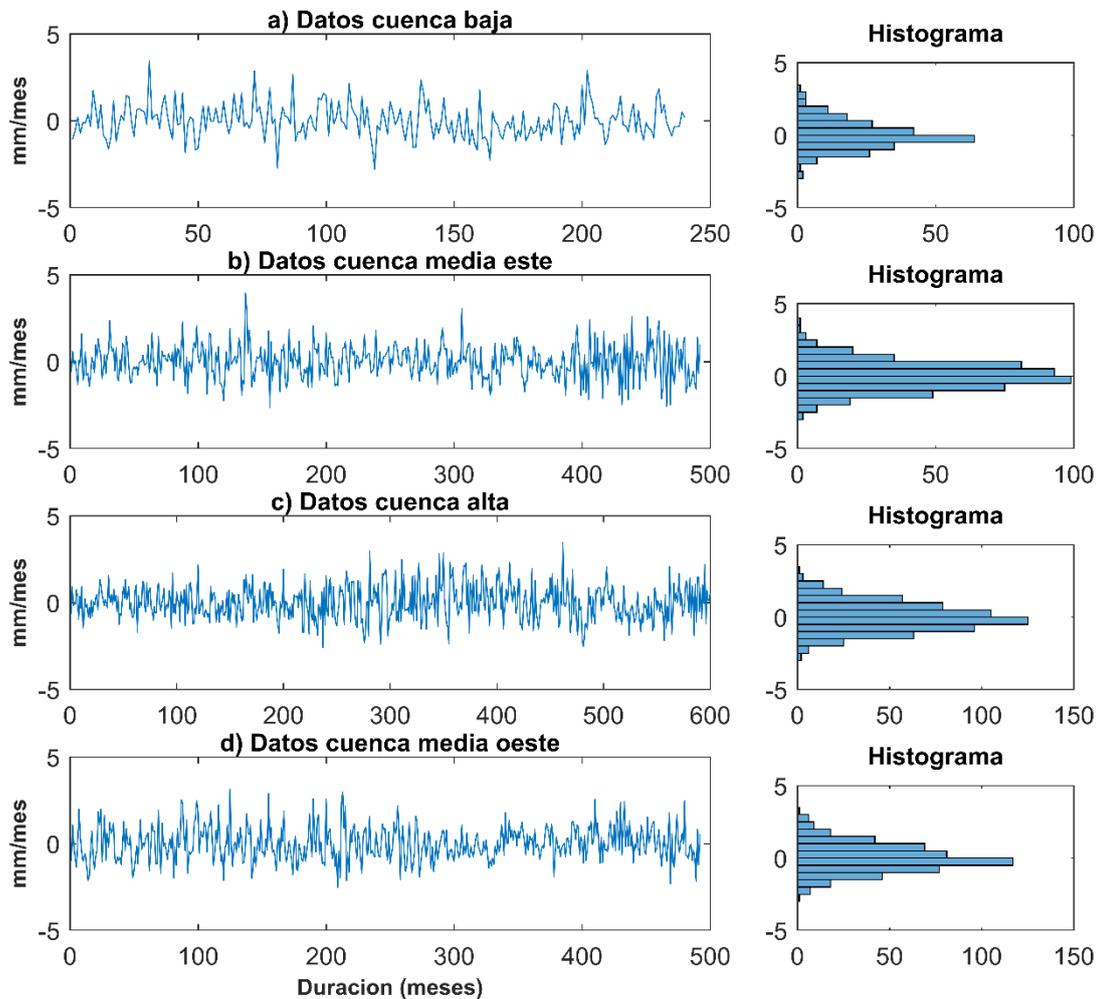


Figura 10: Series estandarizadas.

4.4.2 Cálculo del orden del modelo ARMA(p,q)

El segundo paso fue determinar el orden del modelo ARMA(p,q), sabiendo que el orden queda definido cuando se identifican los valores de p y q, para lo cual se analizó los correlogramas de la FAC y de la FACP, tanto para las series originales y transformadas. Los correlogramas de las cuatro series en estudio se presentan en la Figura 11. Para el análisis se considera que la FAC es un buen estimador para la orden de la media móvil q y la FACP es un buen estimador para el orden autorregresivo p . La Figura 11 (a, c, e, g) corresponde a los correlogramas de la función de

autocorrelación de las cuatro series originales (en color rojo), los cuales presentan un decaimiento y crecimiento cíclico, indicando el comportamiento estacional de la serie, esta estacionalidad se removió cuando se estandarizó las series. Así mismo en la Figura 11 se muestra los correlogramas de la FAC y la FACP de las series estandarizadas (en color azul), estos correlogramas son analizados a continuación para las cuatro series de en estudio.

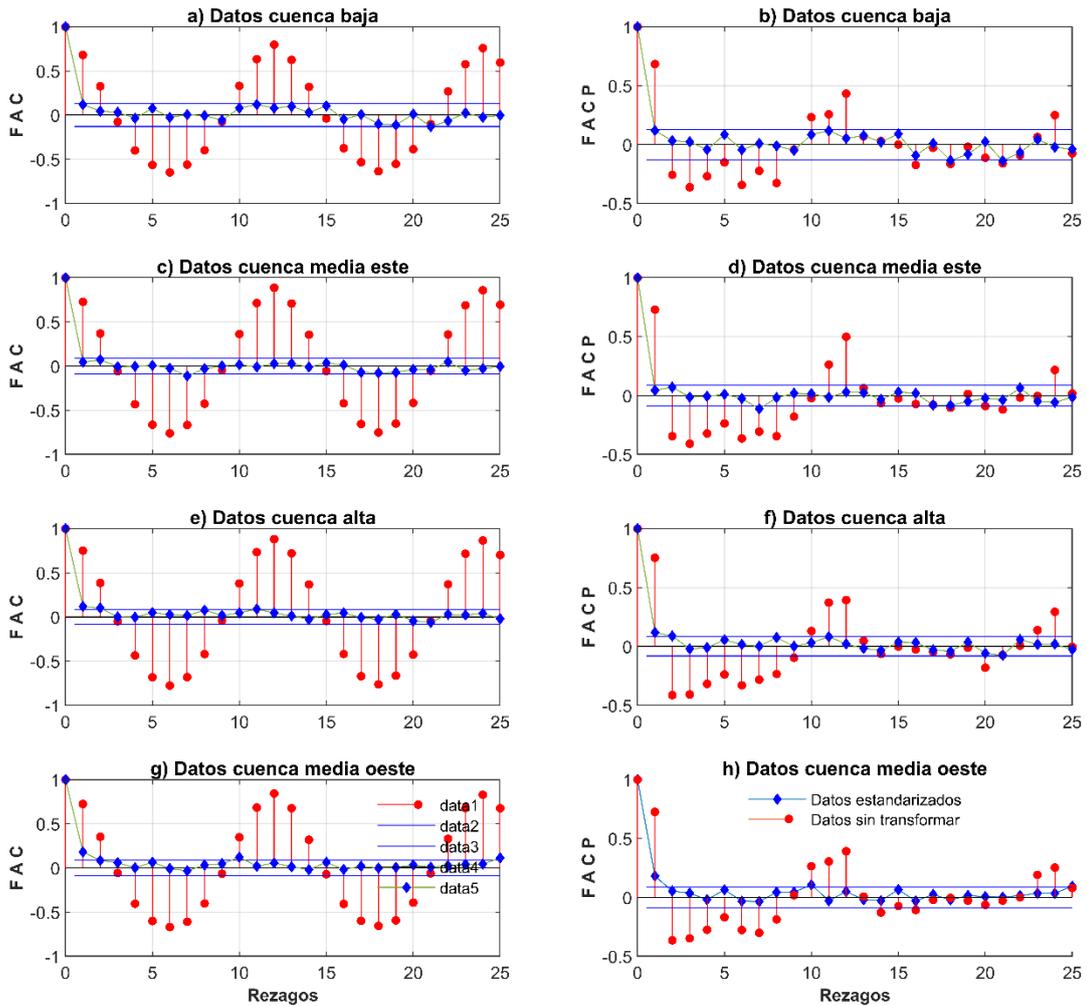


Figura 11: Función de autocorrelación función de autocorrelación parcial

Nota, ACF(a, c, e, g) y PACF (b, d, f, h) para las cuatro series que corresponde a cada región. Para datos originales y estandarizados (Limite de confianza al 95% es igual a $\pm 1.96/\sqrt{N-1}$)

Serie de tiempo estandarizada de la zona homogénea cuenca Baja; La Figura 11a muestra el correlograma de la FAC de la serie estandarizada correspondiente a la cuenca Baja. Se observa que los valores decrecen rápidamente hacia el cero sin llegar

a anularse, presenta el primer valor fuera del límite de confianza en el primer rezago. Caso similar sucede con el correlograma de la FACP, la cual se muestra en el gráfico Figura 11b, la FACP presenta un decrecimiento rápido hacia el cero sin llegar a anularse, su primer valor fuera del límite de confianza acontece en el primer rezago, sin embargo existe otros valores significativos en los rezagos 4, 6, 8 y 9. Indicando que la serie puede ser modelada con un proceso ARMA(1,1).

Serie de tiempo de la zona homogénea cuenca media Este; La Figura 11c muestra el correlograma de la FAC de la serie estandarizada correspondiente a la cuenca media Este, la FAC presenta un decrecimiento rápido con tendencia al cero sin llegar a anularse, sin embargo, presenta un valor fuera del límite de confianza, el cual se encuentra en el rezago siete. Caso similar acontece con el correlograma de la FACP para la misma serie mostrada en la Figura 11d. En la que se observa un decrecimiento rápido de la FACP hacia cero y su primer valor fuera del límite de confianza se encuentra en el séptimo rezago, esto indica que la serie puede ser modelada con un proceso mayor a un modelo ARMA(1,1).

Serie de tiempo de la zona homogénea cuenca Alta; La Figura 11e muestra el correlograma de la FAC de los datos estandarizados de la cuenca Alta, observando un decrecimiento rápido con tendencia al cero, sus dos primeros valores que se encuentran fuera del límite de confianza están en el primer y segundo rezago, De forma similar acontece con el correlograma de la FACP mostrada en la Figura 11f, observando un decrecimiento rápido hacia cero sin llegar a anularse y sus primeros valores fuera del límite de confianza se encuentran en los dos primeros rezagos, esto indicaría que se trata del proceso ARMA(1:2,1:2) o las combinaciones que pueda dar lugar dicho modelos.

Serie de tiempo de la zona homogénea cuenca media Oeste; La Figura 11g presenta el correlograma de la FAC de los datos estandarizados de la cuenca media Oeste, observando un decrecimiento rápido hacia cero sin llegar a anularse y sus primeros valores que se encuentran fuera del límite de confianza están en el rezago uno y dos, para los demás rezagos no son significativos. De igual forma se observa en el correlograma de la FACP presentado en la Figura 11h, donde se observa un decrecimiento rápido hacia el cero sin llegar a anularse y su primer valor fuera del límite

de confianza se encuentran en el rezago 2, siendo para los demás no significativos, indicando que se trata de los procesos ARMA(1,1) o ARMA(2,1)

Finalmente, del análisis de los autocorrelogramas de las FAC y FACP de las cuatro series, se obtiene el orden preliminar de los modelos autorregresivos ARMA. Existiendo en algunos casos más de un modelo propuesto (cuenca media Oeste y cuenca Alta), en otros casos el orden del modelo es alto (cuenca Baja), por estas razones, este análisis no es definitivo. Se hace necesario aplicar otro método para complementar y determinar con mayor rigurosidad el orden del modelo autorregresivo de medias móviles para cada serie.

4.4.3 Determinación del orden del modelo ARMA usando el BIC

La selección del orden de un modelo ARMA(p,q) analizando los correlogramas de las FAC y FACP es bastante desafiante y puede conducirnos al error, seleccionando modelos no parsimoniosos. Sin embargo, el orden de los modelos seleccionados mediante este método se puede tomar como referencia y perfeccionarlos, para lo cual se aplicó el Criterio de Información Bayesiano, el cual se basa en la estimación de la varianza de los parámetros del modelo estimados, este criterio fue definido en la sección 3.8.3.

Para aplicar el BIC se consideró los 16 modelos producto de la combinación de los cuatro primeros modelos autorregresivos; AR(1), AR(2), AR(3) y AR(4). Con los cuatro primeros modelos de media móviles; MA(1), MA(2), MA(3) y MA(4). Dando como resultado 16 modelos ARMA(p,q) los cuales son ARMA(1,1), ARMA(1,2), ARMA(1,3), ARMA(1,4), ARMA(2,1), ARMA(2,2), ARMA(2,3), ARMA(2,4), ARMA(3,1), ARMA(3,2), ARMA(3,3), ARMA(3,4), ARMA(4,1), ARMA(4,2), ARMA(4,3) y ARMA(4,4). Para elegir el modelo entre los propuestos se debe considerar al que obtenga el menor valor BIC entre los 16 modelos, esta evaluación se realizó para las cuatro series en estudio.

En la Figura 12a se muestran los valores del BIC para los 16 modelos propuestos para la serie de tiempo que corresponde a la cuenca Baja. Para esta serie se eligió el modelo ARMA(1,1), por obtener el menor valor BIC (683.9). De la misma forma, se procedió para la serie de tiempo correspondiente a la cuenca media Este, los valores BIC para los 16 modelos evaluados se muestra en la Figura 12b, identificando el menor valor

BIC igual a 1399.3 el que corresponde al modelo ARMA(1,1). Así mismo, En la Figura 12c, se muestran los valores BIC para los 16 modelos evaluados para la serie de tiempo de la zona homogénea cuenca Alta, eligiendo el modelo ARMA(1,1) para esta serie, por obtener el menor valor BIC (1701.2). Finalmente para la serie de tiempo correspondiente a la zona homogénea cuenca media Oeste, se eligió el modelo ARMA(1,1) con valor BIC de 1390.0, por ser el menor valor de los 16 modelos evaluados. En resumen, para las cuatro series evaluadas mediante el análisis del BIC, se obtienen para todas el modelo ARMA(1,1).

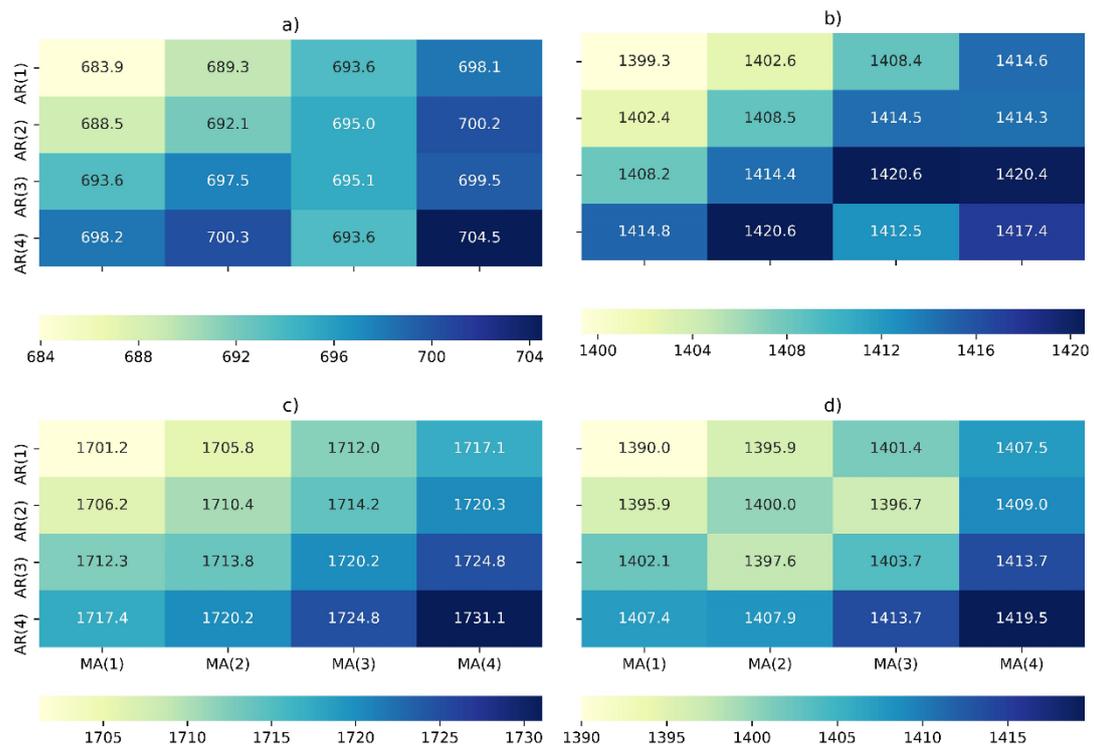


Figura 12: Mapa de colores para la matriz de AR y MA, para valores BIC.

Nota, a) cuenca Baja, b) cuenca media Este, c) cuenca Alta y d) cuenca media Oeste

4.4.4 Ajuste del modelo ARMA y generación de series sintéticas

En este paso se procedió a determinar los parámetros de cada modelo, utilizando el método de máxima verosimilitud a un nivel de significancia de 0.05, los resultados del ajuste se muestran en la Tabla 6. Basados en estos resultados se planteó los modelos autorregresivos ARMA(1,1) para las cuatro series que corresponden a cada zona homogénea identificada para la cuenca del Mantaro. El modelo autorregresivo ARMA(1,1) para la zona homogénea cuenca Baja se formula de la siguiente forma:

$$y_t = 0.00634462 + 0.240456y_{t-1} - 0.168318\varepsilon_{t-1} + \varepsilon_t$$

Del mismo modo, se presentan los modelos para las tres zonas homogéneas restantes; cuenca media Este, cuenca Alta y cuenca media Oeste respectivamente:

$$y_t = 0.00303902 + 0.600837 y_{t-1} - 0.540584 \varepsilon_{t-1} + \varepsilon_t$$

$$y_t = -0.00129037 + 0.382472 y_{t-1} - 0.248941 \varepsilon_{t-1} + \varepsilon_t$$

$$y_t = -0.00129037 + 0.382472 y_{t-1} - 0.248941 \varepsilon_{t-1} + \varepsilon_t$$

Para todas las ecuaciones y_t representa un valor en el momento t de la serie, mientras que y_{t-1} representa la observación inmediatamente anterior y ε_t es el componente aleatorio con distribución normal con media cero y varianza uno.

Tabla 6: Ajuste del modelo ARMA

Región homogénea	Modelos ARMA(p,q)						Validación
	AIC BIC	(p,q)	Int.	VAR	AR	MA	Ljuan –Box (P valor)
Cuenca Baja	683.9	(1,1)	0.00634462	0.9968	0.240456	-0.168318	0.4713
Cuenca media Este	1399.3	(1,1)	0.00303902	0.8750	0.600837	-0.540584	0.3461
Cuenca Alta	1701.2	(1,1)	-0.00129037	0.9596	0.382472	-0.248941	0.2695
Cuenca media Oeste	1390.0	(1,1)	-0.00129037	0.9596	0.382472	-0.248941	0.6903

Nota, p y q son parámetros del modelo, Int es el intercepto, VAR es variación, AR es el componente autorregresivo y MA es componente media móvil.

Los cuatro modelos ARMA(1,1) obtenidos son modelos básicos, parsimoniosos y con ciertas limitaciones cuando se trata de series estacionales. Sin embargo, luego de remover la estacionalidad de las series o al transformar los datos, estos modelos se pueden aplicar a las series en estudio.

4.4.5 Validación del modelo

La validación de los cuatro modelos identificados en cada zona homogénea se realizó comparando los datos generados por cada modelo con los datos observados de precipitación mensual. Para esta validación se escogió el intervalo de tiempo de 20 años, comprendida entre los años 1965 a 1985, la confiabilidad de los datos estimados frente a los datos observados, fue evaluada mediante tres pruebas estadísticas: coeficiente de determinación (R^2), coeficiente de correlación (r) y el error cuadrático medio (RMSE). Los resultados se muestran en la Figura 13.

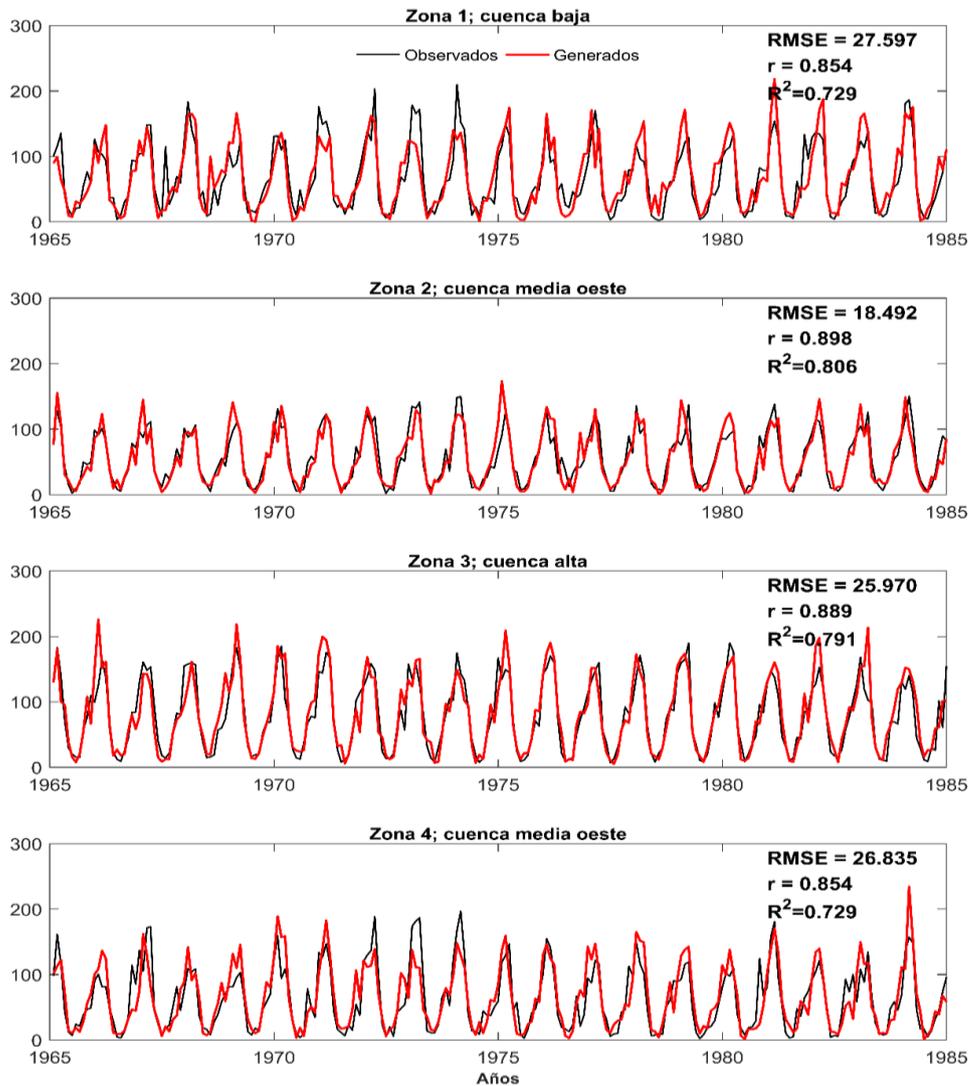


Figura 13: Validación del modelo.

Nota, las Series de precipitación observadas y generadas, RMSE es el error cuadrático medio, r coeficiente de correlación, R^2 coeficiente de determinación.

Para las cuatro series en evaluación, los valores del RMSE son bajos ($RMSE < 27.597$), mientras los coeficientes de determinación y correlación presentan valores altos ($r > 0.854$, $R^2 > 0.729$), indicando una correlación significativa para el nivel de significancia $p > 0.05$. Validando de esta manera los cuatro modelos.

4.4.6 Generación y Validación las series

Se generó 500 años de series sintéticas utilizando las funciones autorregresivas de medias móviles identificadas para cada zona homogénea. Las series generadas se validaron utilizando la prueba de Ljung–Box, cuyos resultados se muestran en la Tabla 6. A un nivel de significancia del 5%, los p -valores de las cuatro series evaluadas son

mayores de 0.05, por lo que la Hipótesis nula no se rechaza. Indicando la independencia de las autocorrelaciones de los residuos. Así mismo, los valores altos obtenidos en esta prueba indican la buena calidad de los modelos. Concluyendo que las cuatro series generadas con los modelos ARMA mantienen las características de las series originales, como su media y varianza.

4.5 ÍNDICE ESTANDARIZADO DE PRECIPITACIÓN (SPI)

Para la caracterización de las sequías en las cuatro regiones identificadas, se determinó el SPI en la escala de 3 meses; para el análisis de las sequías a corto plazo, y a escala de 12 meses; para el análisis de las sequías a largo plazo. Los resultados se presentan en las Figuras 14 y 15, las áreas de color rojo representan valores negativos del SPI, indican una precipitación inferior a la media. Mientras que los de color azul representan precipitaciones superiores a la media, indicando periodos húmedos. Sin embargo, para que un periodo sea considerado como un evento de sequía, los valores del SPI deben ser iguales o menores a -1, considerando al primer mes como el inicio de la sequía, mientras que la duración de la sequía está determinada por los números de meses consecutivos con valores de SPI menores a -1, así mismo el tiempo entre dos eventos de sequía, se contabiliza desde el inicio del evento de sequía hasta el inicio de un nuevo evento, mientras que la severidad del evento se considera como la suma de los valores negativos SPI en un evento y la intensidad es determinado por el menor valor de SPI.

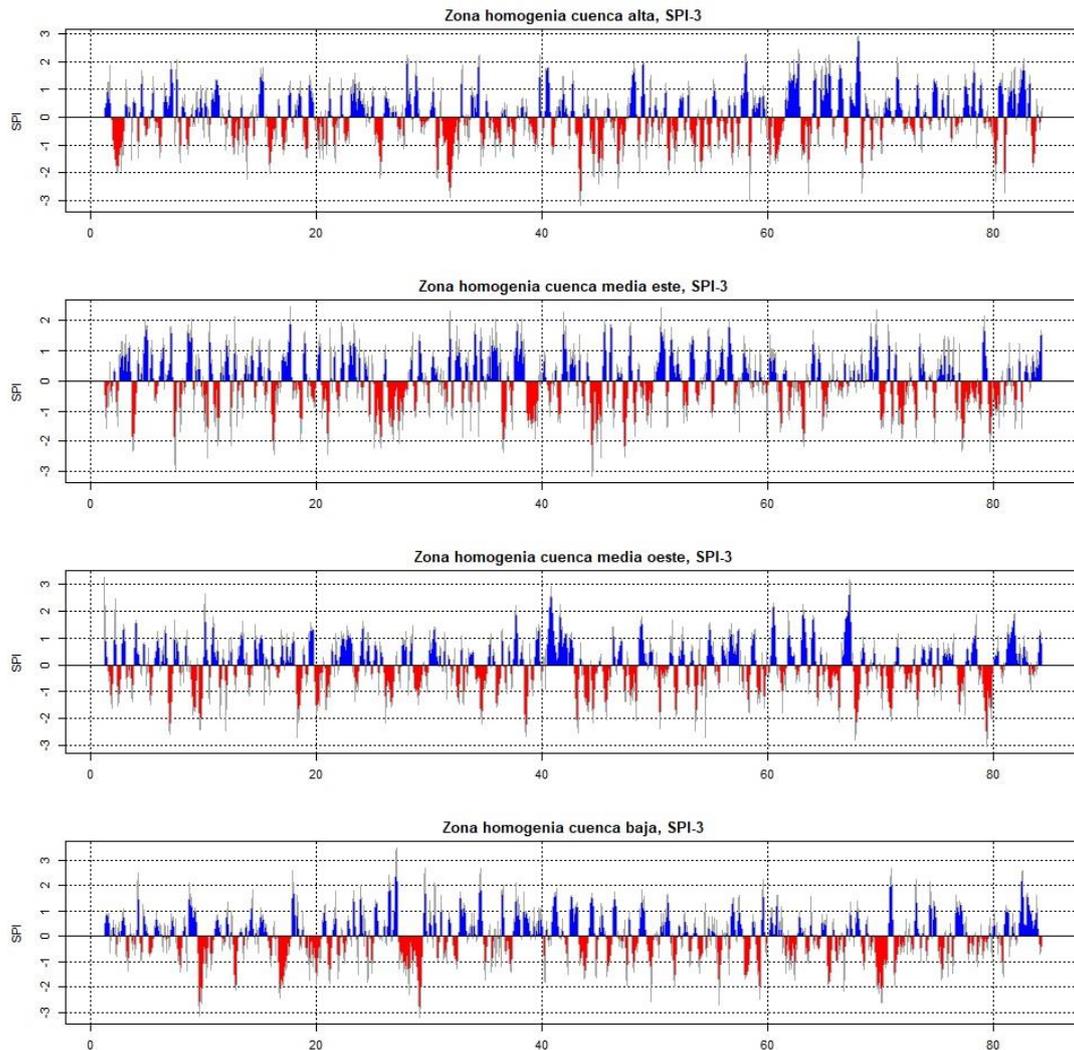


Figura 14: SPI para un intervalo de tiempo de 3 meses.

Nota, se muestra los primeros 80 meses de los 500 años de series sintéticas.

En las Figuras 14 se muestran los SPI en la escala de tiempo de 3 meses para las cuatro series, los eventos identificados tienen una gran variabilidad y se muestran más eventos de corto plazo. Mientras que, para el SPI a escala de 12 meses mostrada en la figura 15, los eventos de sequía ocurren con menos frecuencia, pero son más severas e intensas. En esta escala de tiempo más largo no necesariamente se detectan todas las desviaciones negativas que son evidentes en escalas de tiempo más cortas. También, se evidencia que los valores del SPI-3 muestran un ligero aumento en la precipitación durante los períodos secos, lo que no necesariamente refleja una mejora en la situación de sequía en la escala de tiempo de 12 meses. Al analizar los períodos generados debemos tener en cuenta que varios meses consecutivos de valores de índice negativos no necesariamente indican sequía.

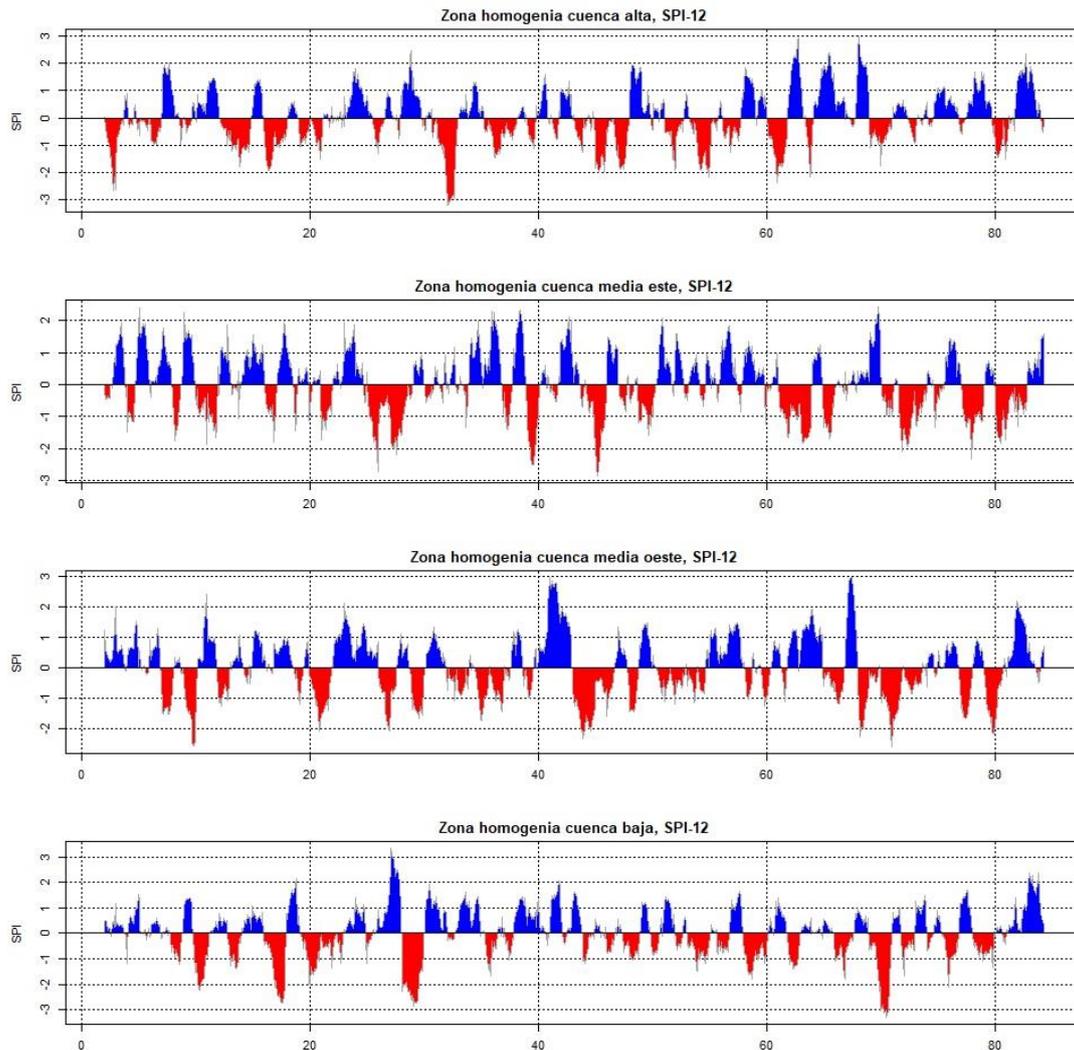


Figura 15: SPI para un intervalo de tiempo de 12 meses

Nota, se muestra los primeros 80 meses de los 500 años de series sintéticas.

4.5.1 Características de las sequías

Para determinar las características de las sequías se analizó las cuatro series SPI, cada uno consistente en 6000 datos. Las características que se identifican son la duración, el tiempo entre sequías, la severidad e intensidad. Así mismo se determina el número de sequías y finalmente se clasifica las sequías en moderadas, severas y extremas, de acuerdo con la Tabla 1. Los resultados de este análisis se muestran en la Tabla 7, para las cuatro zonas en estudio y para los valores de SPI-3 y SPI-12.

Para las sequías a corto plazo (SPI-3) no existe diferencias entre el número de sequías moderadas entre las cuatro zonas homogéneas de la cuenca del Mantaro, sin embargo, para el caso de sequías severas, la cuenca media Este presenta mayor número de

eventos, con 146 eventos identificados. Así mismo, para la cuenca media Oeste se contabilizo el mayor número de eventos extremos identificados. Del total de eventos extremos identificados en las cuatro zonas en estudio, acontecen sequías que alcanzan como máximo 10 meses de duración, con una severidad máxima de 20.15, con una intensidad máxima de 4.49. Teniendo en cuenta, que estos valores no siempre se presentan en un solo evento.

También se observa en la Tabla 7 los resultados para las sequías a largo plazo (SPI-12), por ejemplo, en la cuenca media Este se identificó el mayor número de eventos: 127 moderados, 52 severos y 36 extremos. Del total de eventos identificados en la cuenca, se tiene eventos con una duración máxima de 35 meses, también existe eventos con una severidad máxima de 20.15 e intensidad máxima de 4.49. Considerando que estos valores no siempre se presentan en un solo evento.

Tabla 7: Caracterización de las sequías

Zona	SPI	Tipo de sequías			Total	Valores máximos		
		Moderada	Severa	Extrema		Duración	Severidad	Intensidad
1	3	223	130	93	446	10	20.15	4.48
2	3	222	146	98	466	10	19.21	3.52
3	3	220	127	95	442	10	17.88	3.71
4	3	220	113	102	435	10	18.00	4.49
Promedio		221	129	97	447	10*	20.15*	4.49*
1	12	103	44	31	178	27	52.23	3.41
2	12	127	52	36	215	27	56.12	3.33
3	12	118	46	27	191	23	41.40	3.60
4	12	87	41	34	162	35	53.55	3.30
Promedio		109	46	32	187	35*	56.12*	3.60*

Nota; SPI es índice de precipitación estandarizado, * valores máximos, cuenca Baja (1), cuenca media Este (2), cuenca Alta (3) y cuenca media Oeste (4).

Para las sequías a largo plazo (SPI-12) se tiene en promedio 109 eventos moderados, 46 severos, 32 extremos, haciendo un total de 187 eventos de sequías identificadas en los 500 años en estudio. Estas sequías a largo plazo tienen una relacionan con los caudales de los ríos tributarios del río Mantaro, así como también con el nivel de las represas existentes, sobre todo en la cuenca Alta (Arriaga et al. 2019). Mientras que para las sequías a corto plazo (SPI-3) se obtienen en promedio 221 eventos moderados, 129 severos, 97 extremos haciendo un total de 447 eventos de sequías. La identificación de estas sequías está en relación con la humedad en el suelo (Svoboda

et al. 2012), por lo que su incidencia en la cuenca Este, tiene consecuencias negativas en el sector agrícola.

4.6 SELECCIÓN DE FUNCIONES PROBABILÍSTICAS (PDF)

El análisis univariado de las cuatro variables aleatorias: tiempo entre sequías, duración, severidad e intensidad. Inició eligiendo la función de distribución de probabilidad teórica para cada una de las variables. Para lo cual se ensayaron las FDP más utilizadas en los estudios hidrometeorológicos y en recursos hídricos, tales como Gamma, log-normal de dos parámetros, Gumbel, Pearson Tipo III, Weibull, Exponencial y Pareto. Para determinar la FDP más adecuada para cada variable se utilizó el criterio del menor valor del error cuadrático medio (RMSE), eligiendo a la función con menor valor RMSE: Este criterio fue utilizado en los estudios realizados por Alam et al. (2018) y Chen y Guo (2019), quienes usaron el RMSE para seleccionar las FDP para variables de sequías, obteniendo buenos resultados. Los parámetros de los modelos son estimados por el método de máxima verosimilitud en el intervalo de confianza del 0.95. Los resultados de los ajustes estadísticos se muestran en las Tablas 12, 13, 14 y 15, Anexo 8.7. Como ejemplo se muestran los resultados de la zona homogénea cuenca media Este en la Tabla 8, las letras en negritas se usaron para señalar las funciones de probabilidades seleccionadas para cada variable. Para la serie SPI-3, la FDP para las variables *Tes* y *D*, es la distribución Weibull, por tener el menor valor de RMSE, sus parámetros de localización (β) y escala (α) para el *Tes* es, 11.239 y 2.337 y para *D* es 1.151 y 1.666 respectivamente. Para la variable *S* se seleccionó la FDP Log Normal con parámetro β y α igual a 0.894 y 0.699 respectivamente. Para la variable *I* se obtuvo la FDP Pearson tres parámetros, con β , α y λ igual a 0.894, 0.699 y 1.362 respectivamente.

En la Tabla 8, también se muestra las FDP para el SPI-12 para la misma zona homogénea. Optando para la variable *Tes* la FDP Gamma, con parámetro β y α igual a 0.735 y 31.506 respectivamente. Para la variable *D* y *S* se eligió la FDP Pareto con parámetro β igual a 0.088 y 0.358 respectivamente y parámetros α igual a 4.226 y 4.726 respectivamente. Finalmente, para la variable *I* se tiene la FDP Pearson cuyo parámetro de localización (β) es igual a 0.542, parámetro de escala (α) igual a 0.235 y el parámetro de forma (λ) igual a 1.234.

Tabla 8: Parámetros de las FDP, para la zona 2; cuenca media Este.

Modelo	parámetros	SPI-3				SPI-12			
		<i>Tes</i>	<i>D</i>	<i>S</i>	<i>I</i>	<i>Tes</i>	<i>D</i>	<i>S</i>	<i>I</i>
Gamma	RMSE	0.000	0.012	0.005	0.002	0.002	0.010	0.009	0.005
	β	1.291	2.921	2.098	11.788	0.735	1.137	0.890	11.281
	α	8.332	0.708	1.507	0.138	31.506	4.069	7.943	0.135
Log-normal	RMSE	0.002	0.013	0.004	0.001	0.006	0.008	0.005	0.004
	β	1.941	0.546	0.894	0.440	2.327	1.032	1.297	0.377
	α	1.023	0.583	0.699	0.288	1.517	0.987	1.119	0.290
Gumbel	RMSE	0.017	0.019	0.018	0.009	0.013	0.020	0.026	0.014
	B	16.217	2.833	4.566	1.895	36.819	7.409	12.310	1.799
	A	14.628	1.960	3.779	0.617	32.069	6.705	13.639	0.613
Pearson	RMSE	0.001	0.051	0.004	0.001	0.028	0.071	0.004	0.001
	B	0.380	4.942	1.156	0.192	5.117	3.022	1.910	0.542
	α	5.065	0.228	0.727	0.330	9.855	0.004	1.028	0.235
Weibull	λ	5.575	1.046	1.504	1.362	2.926	1.001	1.523	1.234
	RMSE	0.000	0.011	0.004	0.004	0.003	0.008	0.007	0.007
	β	11.329	2.337	3.508	1.803	20.987	4.674	6.554	1.701
Exponencial	α	1.151	1.666	1.412	3.291	0.824	1.022	0.878	3.128
	RMSE	0.001	0.021	0.009	0.049	0.005	0.008	0.011	0.049
	β	10.758	2.069	3.162	1.621	23.167	4.628	7.066	1.525
Pareto	RMSE	0.001	0.016	0.006	0.031	0.005	0.007	0.005	0.031
	β	-0.089	-0.22	-0.14	-0.676	0.111	0.088	0.358	-0.675
	α	11.697	2.472	3.599	2.387	20.683	4.226	4.726	2.266

Nota: FDA es función densidad de probabilidad, parámetro de localización es β , parámetro de escala es α , parámetro de forma es λ .

4.7 ANÁLISIS UNIVARIADO

El propósito de ajustar una FDP a cada variable de la sequía es realizar el análisis estadístico univariado, además es la base para el análisis bivariado. Puesto que las cópulas están construidas a partir de FDP univariadas. En la Figura 16, se muestran la FDP con mejor ajuste para las variables *Tes*, *D*, *S* e *I*, para las cuatro series SPI-3. Identificando la función Pareto Generalizado y Weibull para la variable *Tes*, también se identificó la función Weibull para la variable *D*, si bien es cierto estas variables son discretas, por lo que les corresponden funciones discretas como la FDP Geométrica (Mathier et al. 1992), sin embargo, por el teorema de Sklar se requiere que las FDP sean continuas para ser usadas en el análisis bivariado (Sklar 1959), por lo que la asignación de la función Pareto Generalizado y Weibull a las variables *Tes* y *D* es apropiada. Finalmente, para las variables *S* e *I* se eligió las FDP Log-normal y Pearson.

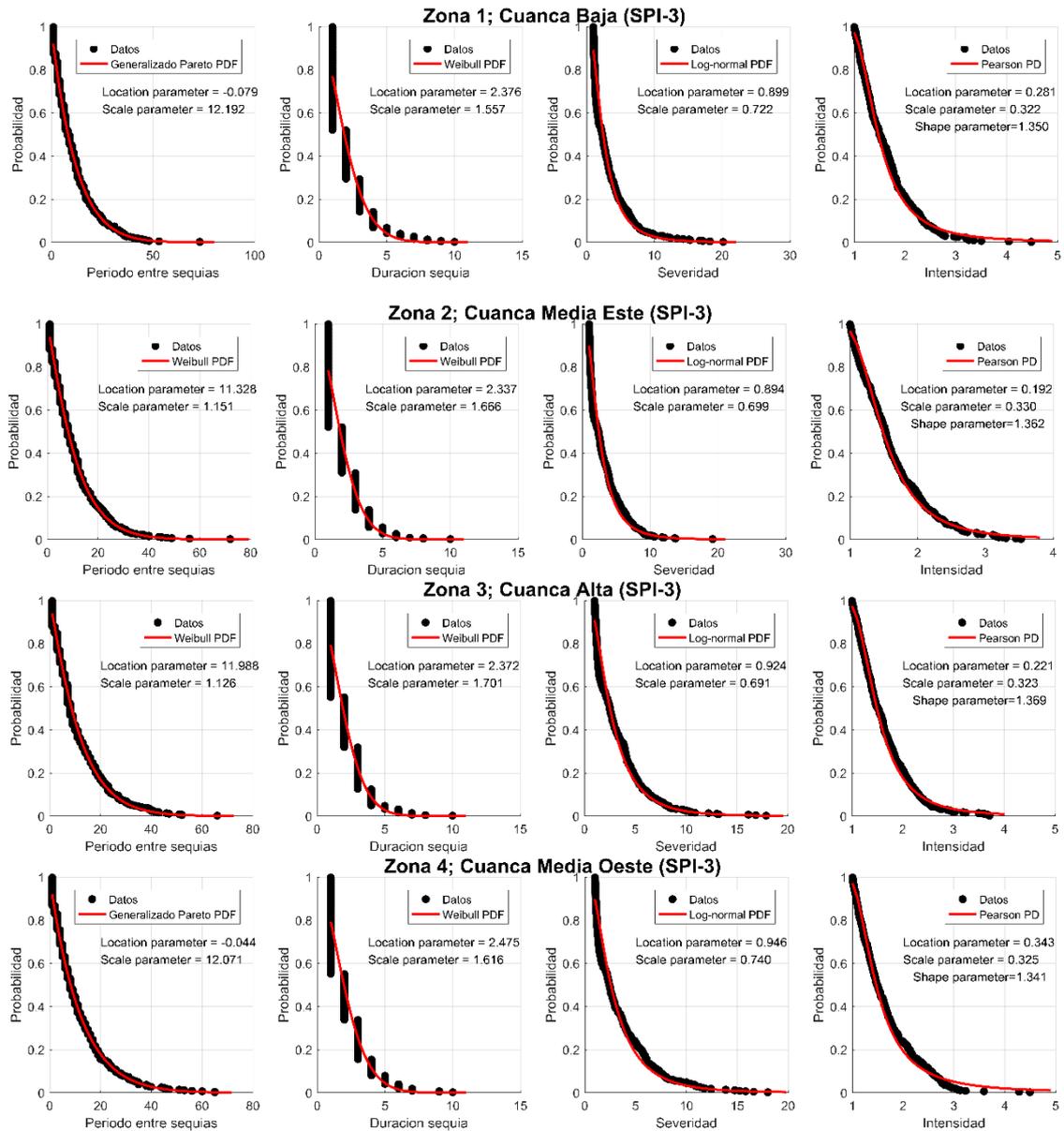


Figura 16: Funciones de probabilidad para ISP 3

Para la variable *Tes* se ha seleccionado las distribuciones de probabilidad Pareto generalizado en la cuenca Baja cuenca media Oeste, Weibull en la cuenca media Este y cuenca Alta. Del mismo modo, para la variable *D* para las cuatro series SPI, se seleccionó la distribución Weibull. La FDP log-normal fue seleccionada para la variable severidad de sequías, en las cuatro zonas homogéneas en estudio y finalmente para la variable intensidad se ha determinado FDP Pearson en las cuatro regiones homogéneas identificadas en la cuenca del Mantaro, la selección de estos modelos obedece al criterio del menor valor del error cuadrático medio, presentados en la Tabla 8.

Así mismo en la Figura 17 se muestran la FDP con mejor ajuste para las variables *Tes*, *D*, *S* e *I*, para las cuatro series correspondientes a la sequía a largo plazo (SPI-12). Para la variable *Tes* se ha seleccionado las distribuciones de probabilidad Gamma, en la cuenca Baja, cuenca media Este y cuenca media Oeste; Weibull, para la cuenca Alta. Del mismo modo se han seleccionado para las cuatro series las distribuciones la función Pareto para la *D* y para la *S*; Pearson, para la variable *I*.

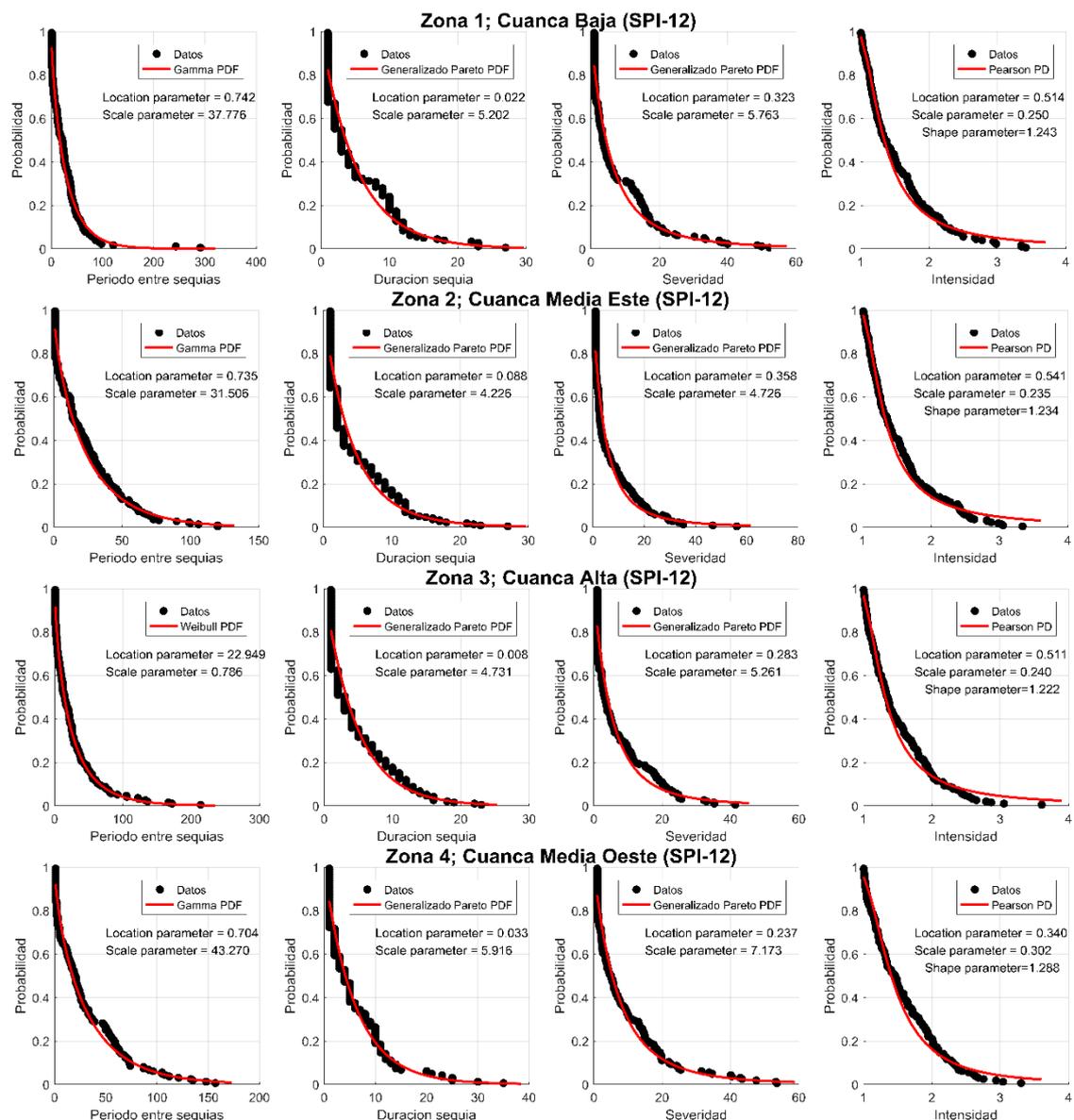


Figura 17: Funciones de probabilidad para SPI-12

4.8 ANÁLISIS BIVARIADO

Las sequías son fenómenos complejos su evaluación en términos de una variable no proporciona un análisis exhaustivo, por lo que, el análisis por separado de la

distribución de la duración y severidad de la sequía no puede revelar la correlación significativa entre ellas. Shiau y Modarres (2009) sugieren que, a cambio de utilizar el análisis univariado tradicional para la evaluación de la sequía se debe usar el análisis bivariada, el cual se basa en el análisis de correlación, que para el caso de las sequías se estudia las correlaciones entre la severidad, duración, tiempo entre sequías e intensidad. Así mismo, Mishra y Singh (2010) reafirman que el análisis bivariado, basado en la distribución conjunta, es el enfoque adecuado para describir las características correlacionadas de la sequía.

4.8.1 Análisis de correlación

La Figura 18, muestra los gráficos de dispersión entre las cuatro variables en estudio; *Tes*, *D*, *S* e *I*. Para las sequías a corto plazo (SPI-3) y largo plazo (SPI-12), para las cuatro zonas homogéneas identificadas. También, se muestran los valores de los coeficientes de correlación de Pearson (*P*); Rho de Spearman (*S*) y Thao de Kendall (*K*). Los resultados muestran valores negativos para las relaciones entre la variable *Tes* y las otras variables *D*, *S* e *I*, en las cuatro zonas homogéneas evaluadas y para el SPI-3 y SPI-12, indicando una relación inversa, excepto en la zona 4 para el SPI-3. Sin embargo, los valores obtenidos son cercanos al cero para las tres relaciones, indicando, que la variable *Tes* no tienen una relación significativa con las otras variables.

La correlación bivariada entre la severidad y la duración para las series SPI-3 y SPI-12, en las cuatro zonas homogéneas es alta. La correlación positiva indica la relación directa entre estas dos variables, siendo los valores de *r* de Pearson y rho de Spearman mayores de 0.9. Teniendo en cuenta que la correlación se considera alta, cuando estas son mayores a 0,8. Sobre esta correlación alta existente Montaseri et al. (2018) menciona que es la más importante, debido a que se consideran como parámetros primarios en la estimación de otras características de la sequía, como la intensidad. También es considera como característica básica en el monitoreo de las sequías en tiempo real y gestión de la sequía a largo plazo.

Así mismo, la relación entre las variables *D* e *I*, es positiva mostrando una relación directa. Sin embargo, los valores de *r* de Pearson para las series de SPI-12 son mayores a los mostrados por la serie SPI-3 en las cuatro zonas homogéneas. Esta relación que

existe entre estas dos variables es usada para describir los picos de los eventos de sequías.

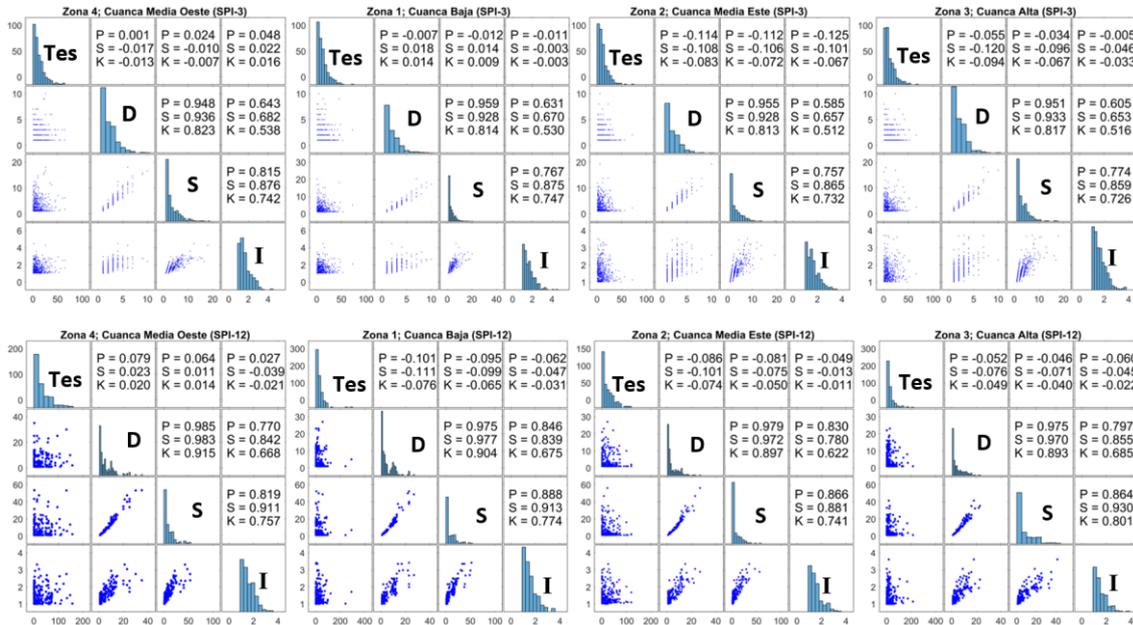


Figura 18: Grafico de dispersión entre las variables de las sequías.

Nota; P es coeficientes de correlación de R de Pearson, S es Rho de Spearman, K es Thao de Kendall, Ter es periodo entre sequías, D es duración de sequías, I es la Intensidad y S es la Severidad de la sequía.

4.8.2 Análisis de cópulas

La alta correlación positiva entre las variables *D* y *S* de las sequías permite el análisis bivariado. Este análisis de frecuencia correlacionando se usa para considerar la ocurrencia y el efecto simultáneo de las dos características. En consecuencia, los análisis de probabilidad conjunta o bivariada de tales características de sequía pueden proporcionar criterios complementarios para desarrollar políticas que optimicen la toma de decisiones en el ámbito de la gestión de los recursos hídricos.

Dado que la duración y la severidad de la sequía se ajustaron a distribuciones de probabilidad diferentes, no se pueden aplicar las distribuciones bivariadas de uso común, como la Exponencial bivariada o la distribución Gamma bivariada. Por lo que en el presente trabajo se emplearon cópulas para vincular las distribuciones univariadas, dando como resultado la distribución bivariada para la relación entre la duración y severidad de la sequía. Estas distribuciones conjuntas o bivariadas son conocidas como cópulas Arquimedianas, en este trabajo consideramos las tres más

usadas; Clayton, Gumbel y Frank (Chen y Guo 2019). Estas utilizan las funciones de distribución univariada determinadas en la sección anterior para componer la estructura de la cópula, mientras que su parámetro se calcula en función al valor de la correlación de Spearman, en algunas cópulas donde no existe solución analítica, la aproximación se basa en un ajuste suave a los valores calculados en valores discretos de los parámetros de la cópula.

Tabla 9: Ajuste para las cópulas Arquimedianas (duración severidad)

ZONA	SPI	Copula arquimedianas	Parámetros	Log-likelihood	AIC
1	3	Clayton	8.779	439.37	-876.73
		Gumbel	4.07	417.68	-833.36
		Frank	14.94	392.78	-783.57
1	12	Clayton	8.7	228.86	-455.71
		Gumbel	6.11	231.91	-461.83
		Frank	40.07	254.33	-506.67
2	3	Clayton	8.69	423.68	-845.35
		Gumbel	3.89	416.14	-830.28
		Frank	14.9	409.19	-816.38
2	12	Clayton	17.35	314.89	-627.77
		Gumbel	6.51	289.23	-576.46
		Frank	26.13	286.33	-570.66
3	3	Clayton	4.9	399.49	-796.97
		Gumbel	3.94	401.4	-800.79
		Frank	6.036	401.49	-800.98
3	12	Clayton	7.37	220.07	-438.14
		Gumbel	5.35	224.36	-446.72
		Frank	10.76	250.26	-498.53
4	3	Clayton	5.02	398.98	-795.96
		Gumbel	4	400.49	-798.99
		Frank	20.80	404.59	-807.18
4	12	Clayton	21.46	255.15	-508.29
		Gumbel	7.14	237.16	-472.31
		Frank	33.11	253.07	-504.14

Nota; Los mejores estimadores se muestran en negrita.

Para determinar la cópula arquimediana que mejor ajuste tiene a los datos bivariados entre duración y severidad, se aplicó la prueba estadística del logaritmo natural de la verosimilitud (Log-likelihood) y el criterio de información Aikaike (AIC) (Ayantobo et al. 2018). Eligiendo la copula con mayor valor en la prueba Log-likelihood y menor valor en AIC, los resultados de estas pruebas se muestran en la Tabla 9, para las ocho series en estudio.

Por ejemplo, para la relación de las variables, duración y severidad, correspondientes a la zona homogénea cuenca Baja, para la serie SPI-3, se eligió la cópula Clayton con parámetro 8.779, por ostentar el mayor valor de Log-likelihood (439.37) y el menor valor AIC (-876.73), de igual forma para la misma zona y para el SPI-12, se selecciona la cópula Frank con parámetro 40.07, por tener el mayor valor de Log-likelihood (254.33) y el menor valor AIC (-506.67), sin embargo de las tres copulas probadas la copula Gumbel no fue considerada en ninguno de los casos.

4.9 PERIODO DE RETORNO BIVARIADO PARA EL SPI-3

El período de retorno bivariado entre las variables D y S , se definió para dos casos: primero, para el período de retorno conjuntivo definido por la condición, $D \geq d$ y $S \geq s$; segundo, para el período de retorno disyuntivo definido por la condición, $D \geq d$ o $S \geq s$. Ambas definidas en la sección 2.13.2. Los resultados de este análisis se muestran en la Figura 19, para las 4 series SPI-3, que corresponde a cada zona homogénea identificada. La Figura 19a corresponde a la función de distribución acumulada bivariada (FDAB) de la relación entre la duración y severidad, con probabilidades de 0.3, 0.5, 0.7 y 0.9. La Figura 19b muestra el periodo de retorno bivariado disyuntivo ($D \geq d$ o $S \geq s$), como líneas de contorno a intervalos de 10, 20, 50, 100 y 200 años. La Figura 19c, muestra el periodo de retorno bivariado conjuntivo ($D \geq d$ y $S \geq s$), como líneas de contorno a intervalos de 10, 20, 50, 100 y 200 años. Para las Figuras 19 y 20, los círculos representan los puntos de datos bivariados, duración y severidad; los círculos de color magenta indican eventos de sequías moderadas; los verdes indican eventos de sequías severas y los rojos representan eventos extremos.

En la Figura 19 se muestra los resultados del periodo de retorno bivariado para la duración y severidad de las sequías para la zona homogénea cuenca Baja y para sequías a corto plazo (ISP-3). Observándose que existe dos eventos, cuya duración es mayor de 7 días o con una severidad mayor de 16 ($D \geq 7$ o $S \geq 16$), que sobrepasan el periodo de retorno de 200 años. También, se identifican 3 eventos bivariados conjuntivos definidos por; $D \geq 7$ y $S \geq 16$. Y 13 eventos de sequía extrema que exceden los 200 años de periodo de retorno, indicando que eventos de sequías que tienen una duración mayor de 6 meses y una severidad mayor de 10, su ocurrencia es menos probable.

El periodo de retorno bivariado disyuntivo de los eventos moderados es menor de 10 años en la cuenca Baja y media, a diferencia de la cuenca Alta, cuyo valor es de 20 años. Caso similar ocurre con el periodo de retorno bivariado conjuntivo para los eventos moderados. Siendo el periodo retorno de estos eventos en la cuenca Baja y Media menor de 50 años, mientras que el periodo de retorno para la cuenca Alta es menor de 200 años, por lo que los eventos moderados con un nivel de severidad y/o duración determinada, son más recurrentes en la cuenca media y Baja, a diferencia de la cuenca Alta.

Para el caso de las sequías a corto plazo (SPI-3), los eventos moderados tienen su periodo de retorno disyuntivo menor a 10 años en la cuenca Baja, media Oeste y media Este, a diferencia de la cuenca Alta, cuyo periodo de retorno es menor a 20 años, por lo que existe diferencia entre la ocurrencia de los eventos moderados en el ámbito de la cuenca del Mantaro. Sin embargo, no existe diferencias entre las cuatro zonas, para los periodos de ocurrencia conjuntivos para el caso de las sequías extremas, indicando que los eventos extremos identificados tienen una ocurrencia mayor de 200 años.

El periodo de retorno disyuntivo ($D \geq d$ o $S \geq s$), para cualquiera de las tres categorías de sequías: moderada, severa y extrema, es menor con respecto al periodo de retorno conjuntivo ($D \geq d$ y $S \geq s$), debido a que los eventos deben cumplir con una de las dos condiciones, tal es el caso que acontece en la cuenca media Este, todos los eventos moderados tienen un periodo de retorno disyuntivo menor de 10 años, mientras su periodo de retorno conjuntivo es menor de 50 años. Indicando que los eventos que presentan una de las dos condiciones son más ocurrentes de aquellos eventos que presentan las dos condiciones juntas en la cuenca.

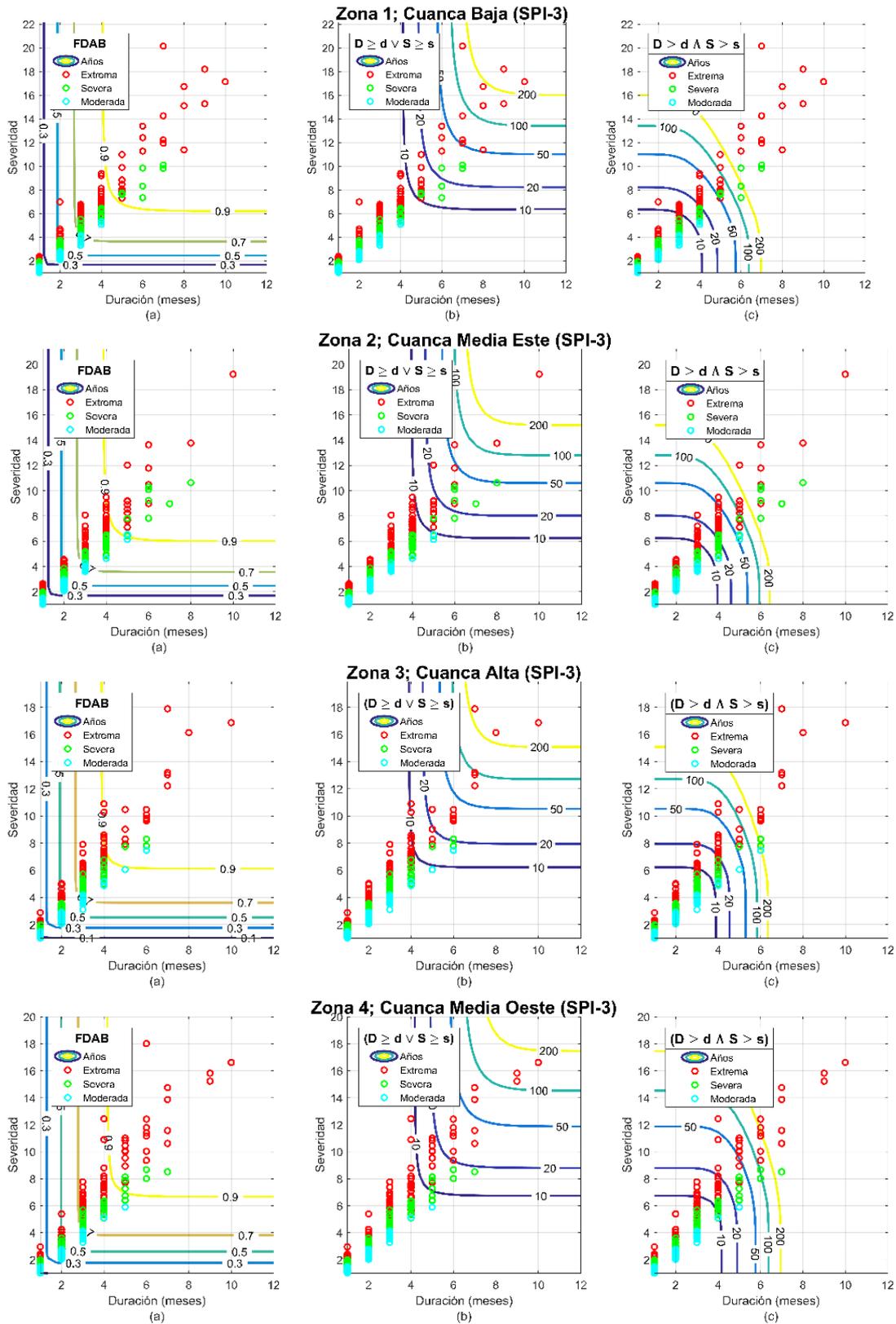


Figura 19: Periodo de retorno bivariado SPI-3.

Para comprender y compara los resultados de los periodos de retorno univariado y bivariado se elaboró la Tabla 10. El periodo de retorno univariado para un evento de 6

meses de duración, es de 65 años. Mientras que, el periodo de retorno univariado para un evento con severidad de 11, es de 50 años. Sin embargo, el periodo de retorno bivariado disyuntivo de $D \geq 6$ o $S \geq 11$, es de 30 años, El periodo de retorno bivariado conjuntivo $D \geq 6$ y $S \geq 11$, es de 400 años. Esto implica que, si se ignora la estrecha correlación entre severidad y duración, empleando un análisis univariado, la ocurrencia del evento de sequía se subestimarán, con respecto a la ocurrencia bivariada conjuntiva, mientras que se sobrevalorará el periodo de ocurrencia con respecto al periodo de retorno disyuntivo. Toda vez que un evento de 6 meses de duración y una severidad 11, tienen una probabilidad de ocurrencia de 400 años, a diferencia del periodo de ocurrencia de un evento de 6 meses de duración o una severidad de 11, que tiene un periodo de ocurrencia de 30 años.

Tabla 10: Periodo de retorno univariado y bivariado, SPI-3.

Duración (meses)	Severidad	Td (años)	Ts (años)	T _{SyD} (años)	T _{SoD} (años)
3	4	4	4	5	3
4	7	9	13	22	7
6	11	65	50	400	30
7	14	204	118	2757	77
9	17	2683	254	75441	233
10	20	11078	510	618520	488

4.10 PERIODO DE RETORNO BIVARIADO PARA UN SPI-12

En la Figura 20 se muestra el periodo de retorno bivariado disyuntivo, conjuntivo y la FDAB de la relación entre las variables duración y severidad, para las cuatro series en estudio con respecto a las sequías a largo plazo (ISP-12). En el caso de la cuenca media Este, para el periodo de retorno bivariado disyuntivo ($D \geq d$ o $S \geq s$), se observa que existe un evento con mayor de 25 días de duración o con una severidad mayor de 55 ($D \geq 25$ o $S \geq 55$), cuyo periodo de retorno sobre pasa los 100 años. De igual forma, se muestra los periodos de retornos conjuntivos ($D \geq d$ y $S \geq s$), identificándose un evento de sequía severa y cuatro eventos de sequía extrema, cuyo periodo de retorno es mayor de 200 años. Indicando, que la ocurrencia de los eventos de sequía que tienen una duración mayor de 20 meses y una severidad mayor de 35, son menos probable.

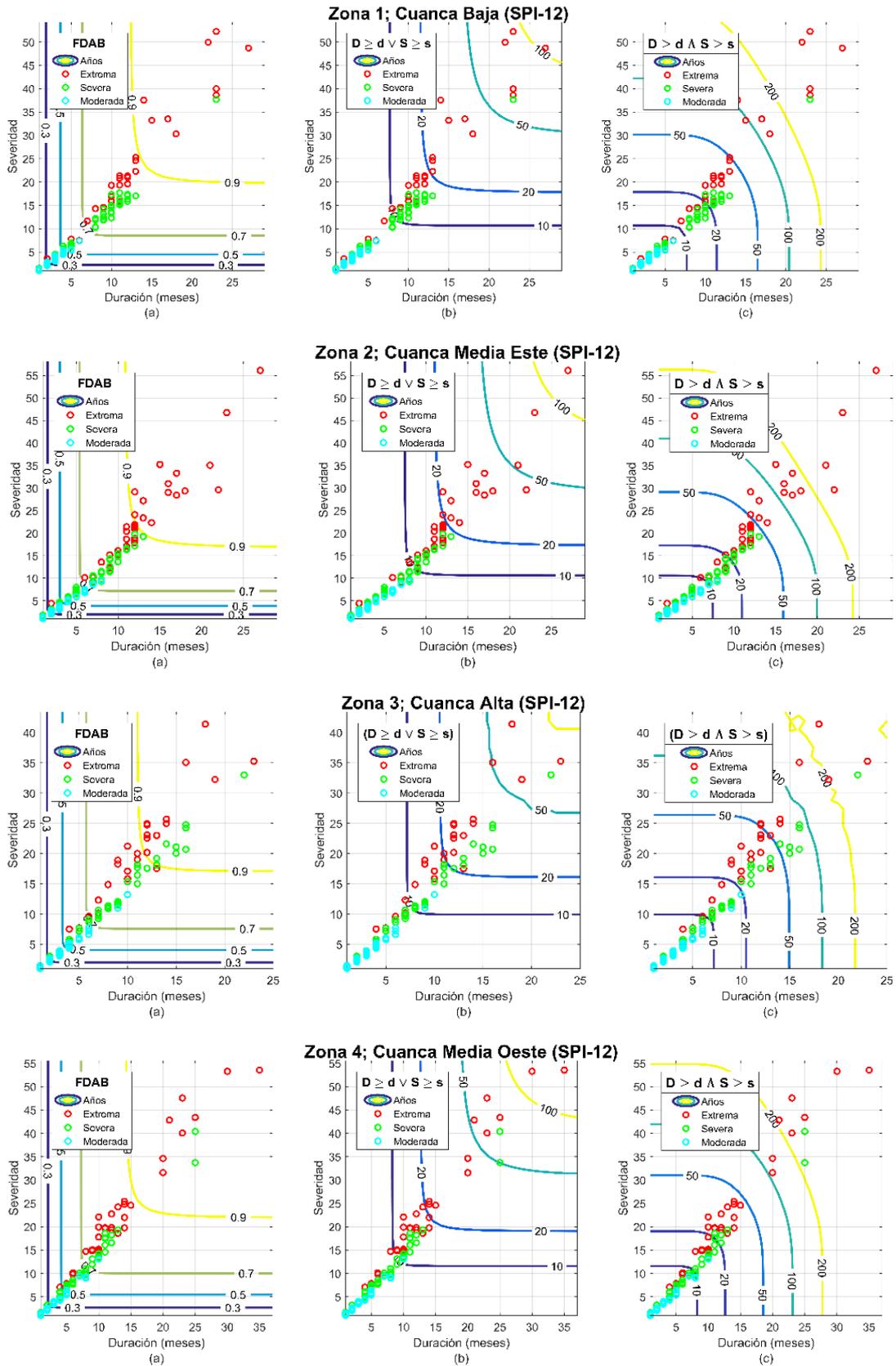


Figura 20: Periodo de retorno bivariado SPI-12.

Para la zona homogénea cuenca media Este, mostrada en la Figura 20, se tiene un evento cuyo periodo de retorno es de 100 años. Debido a que el evento presenta una de las dos condiciones para ser considerado un suceso de éxito; su duración es mayor de 25 o su severidad es mayor de 47. Para el periodo de retorno conjuntivo $D \geq 25$ y $S \geq 47$, para la misma zona homogénea, el mismo evento tienen un periodo de retorno mayor de 200 años. Debido a que, el evento de sequía acontecerá si ambas condiciones se presentan.

En general, se observa en el periodo de retorno disyuntivo, para los eventos categorizados como sequías moderadas, tienen un periodo de retorno menor a 10 años, en la cuenca Baja, media Este y media Oeste, a diferencia de la cuenca Alta cuyo periodo de retorno de estos eventos es menor de 20 años. Caso similar ocurre para los eventos severos cuando se considera el periodo de retorno bivariado conjuntivo, indicando que, para la cuenca Baja, cuenca media Este y Oeste, estos eventos tienen un periodo de recurrencia menor de 20 años, a diferencia de la cuenca Alta cuyo periodo de ocurrencia es mayor de 50 años. Sin embargo, no existe diferencias significativas cuando se trata de eventos extremos, tanto para el periodo de retorno disyuntivo (> 100 años) y periodos de retorno conjuntivos (>200 años).

En la Tabla 11 se observa que el periodo de retorno univariado y bivariado, para las sequías a largo plazo (SPI-12), por ejemplo, para un evento de 23 meses de duración su periodo de retorno es de 159 años, así mismo, el periodo de retorno univariado para una severidad de 44 es de 110 años. Sin embargo, el periodo de retorno disyuntivo para $D = 23$ o $S = 44$ es de 74 años, del mismo modo el periodo de retorno conjuntivo $D = 23$ y $S = 44$ es de 499 años. Indicando que, si se ignora la estrecha correlación entre severidad y duración, empleando un período de retorno univariado, el periodo de retorno del evento de sequía se subestimarán con relación al periodo de retorno bivariado conjuntivo, lo contrario ocurre cuando se compara con el periodo de retorno bivariados disyuntivo.

Tabla 11; Periodo de retorno univariado y bivariado, SPI-12.

Duración (meses)	Severidad	Td (años)	Ts (años)	TSyD (años)	TSoD (años)
5	10	6	9	9	6
10	18	15	20	23	14
14	27	32	41	61	25
18	35	66	67	156	42
23	44	159	110	499	74
27	52	317	160	1316	116

Estos patrones se repiten para las otras zonas homogéneas identificadas en la cuenca del Mantaro (Anexo 8.4). Sin embargo, existe una ligera diferencia entre sequías a corto plazo (SPI-3) y largo plazo (SPI-12).

Los resultados obtenidos en el presente trabajo concuerdan con los obtenidos por Gálvez (2017) en su tesis titulada “Caracterización de las sequías históricas y proyectadas bajo escenarios de cambio climático en la cuenca del río Mantaro” en la que analiza las series observada de precipitación mensual de 37 estaciones (1970-2010) y las precipitación proyectada (2030-2070), sobre la base de seis escenarios de cambio climático, concluyendo que las eventos severos serían más recurrentes en la cuenca media, mientras que las sequias externas tendrían un comportamiento igual en toda la cuenca. En conclusión, modelar un sistema de pronóstico de sequías es uno de los desafíos más importantes en la cuenca del Mantaro, por su incidencia negativa en la agricultura y los recursos hídricos. Este fenómeno se puede estudiar desde la complejidad física de un fenómeno natural (Basu et al. 2017) o desde modelos basados en datos a partir de los patrones observados, sin la necesidad de una comprensión profunda del proceso físico que rodea a la ocurrencia de un evento (Anshuka et al. 2019).

V. CONCLUSIONES

Objetivo específico 1: Zonificación de la cuenca y cálculo de las series tiempo de precipitación por zona.

Se identificó en la cuenca del Mantaro cuatro zonas climáticamente homogéneas; cuenca Baja, cuenca media Este, cuenca Alta y cuenca media Oeste. Para cada zona se obtuvo una serie de precipitación mensual representativa, resultado de los promedios de las series de las estaciones ubicadas en cada zona. Para la cuenca Alta se obtuvo una serie con 600 datos, para la cuenca media Este y Oeste 492 datos y para la cuenca Baja 240 datos. Las cuatro series no presentan tendencias ($M-K > 0.105$) y son homogéneas ($p > 0.51$).

Objetivo específico 2: Generar 500 años de datos de precipitación mensual aplicando modelos autorregresivos de medias móviles.

Del análisis de los correlogramas de las FAC y FACP se concluye que existe una autocorrelación al menos en un rezago en cada uno de las series, por lo que se justifica el uso de los modelos ARMA en el modelamiento y generación de datos. El orden del modelo autorregresivo de medias móviles se identificó mediante el criterio de información bayesiano, siendo el modelo con mejor ajuste para las cuatro series de precipitación mensual el modelo ARMA(1,1). Mientras que, los valores de los parámetros de los modelos fueron determinados utilizando el método de máxima verosimilitud, dando como resultado cuatro funciones autorregresivas, correspondientes a las cuatro zonas climáticas homogéneas. Al comparar las cuatro series originales (1965 - 1985) con los 20 años de serie de precipitación generadas por sus respectivos modelos, se obtuvieron réplicas bastante buenas ($RMSE < 27.59$; $R^2 > 72.9\%$; $r > 85.4\%$), indicando la buena performance de los modelos, constituyéndose una alternativa apropiada para representar las cuatro series históricas. Al evaluar las series mensuales de precipitación generadas (500 años) por medio de la prueba Ljung-Box indican que las cuatro series son válidas ($p > 0.05$). Por lo tanto, las cuatro series

sintéticas reproducen apropiadamente la media y la varianza de las series originales, también estas series proporcionan valores extremos que permitan evaluar circunstancias más inusuales, por consiguiente, las series generadas son adecuadas para ser utilizadas en el análisis de sequías.

Objetivo específico 3: Análisis de correlación entre las variables de las sequías.

Los resultados del análisis de correlación, tanto para las variables de las sequías a largo plazo (SPI-12) y corto plazo (SPI-3), confirman la asociación altamente positiva ($r > 90\%$) entre la duración y severidad. Por lo tanto, es factible el análisis bivariado usando cópulas Arquimedianas.

Objetivo específico 4: Análisis del periodo de retorno univariado y bivariado.

Del análisis comparativo entre los periodos de retorno univariado y bivariados, tanto para las sequías de larga duración (SPI-12) y corta duración (SPI-3), se concluye que acontece una sobrevaloración del periodo de retorno univariado con respecto al periodo de retorno bivariado disyuntivo, mientras que existe una infravaloración del periodo de retorno univariado con respecto a periodo de retorno bivariado conjuntivo. Estos resultados demuestran que el análisis univariado no es suficientes para evaluar la compleja relación que existe entre las variables de las sequías, siendo necesario el análisis bivariado entre la relación de la duración y severidad.

Objetivo específico 5: Determinar el periodo de retorno bivariado para la duración y severidad.

Para las sequías a corto plazo (SPI 3), se encontró que los eventos moderados tienen un periodo de retorno bivariado disyuntivo menor de 10 años en la cuenca Baja, cuenca media Este y cuenca media Oeste, sin embargo, para la cuenca Alta los eventos moderados tienen un periodo de retorno menor de 20 años, por lo que se concluye que los eventos de sequía moderados son más ocurrentes en la cuenca Baja y media a diferencia de la Cuenca Alta. Así mismo, para el periodo de retorno bivariado conjuntivo, los eventos moderados tienen un periodo de retorno menor de 20 años en la cuenca Baja, mientras que para la cuenca media y Alta su periodo de retorno es menor 50 años. Concluyendo que la cuenca Baja es afectada más con eventos moderados con una duración y severidad dada. Sin embargo, para los eventos extremos se tiene un periodo de retorno bivariado disyuntivo menos de 250 años, para las cuatro

zonas, no existiendo diferencias. Caso similar sucede para el periodo de retorno conjuntivo, cuyo periodo de retorno sobre pasa los 200 años. Concluyendo que cuando se trata de sequías extremas, la ocurrencia de estos eventos afecta toda la cuenca.

Para los eventos severos correspondientes a las sequías a largo plazo (SPI 12), se encontró que el periodo de retorno bivariado disyuntivo es menor de 20 años, para la cuenca Baja y cuenca media Este, sin embargo, la ocurrencia de estos eventos es más amplia para la cuenca media Oeste y cuenca Alta, sugiriendo que, los eventos severos acontecen con más frecuencia en la cuenca baja y media este. Con relación a los eventos extremos, los periodos de retorno conjuntivo y bivariado, no tienen diferencias entre las cuatro zonas en estudio, concluyendo que estos eventos son más extensos abarcando toda la cuenca.

VI. RECOMENDACIONES

A continuación, se dan algunas recomendaciones para futuras investigaciones afines.

- Se recomienda realizar estudios que incluyan las variables duración e intensidad, duración y tiempo entre sequías. En general entre las otras variables, para identificar patrones de comportamiento.
- El SPI se basa principalmente en datos de precipitación, por lo tanto, para analizar los tipos individuales de sequía con más detalle, se recomienda usar otros instrumentos o índices de sequía que incluyen otras variables.
- Se recomienda extender el análisis bivariado al estudio multivariante con cópulas dos parámetros, así como incluir en el análisis estadística bayesiana, que incluya los aspectos temporales y espaciales. Además, ampliar las cópulas, como las elípticas; normal y student t. Así como también las cópulas de valores extremos; Husler-Reiss, Galambos, Tawn, Plakett y Farlie.

VII. REFERENCIAS BIBLIOGRÁFICAS

Abramowitz, M. and Stegun, I. 1965. Handbook of mathematical functions, with formulas, graphs, and mathematical tables. 1ra ed. New York, Dover Publications. 1046 p.

Alam, M.; Emura, K.; Farnham, C. and Yuan., J. 2018. Best-fit probability distributions and return periods for maximum monthly rainfall in Bangladesh. *Climate* 6(1):9.

AMS (American Meteorological Society). 1997. Meteorological drought-policy statement. *Bulletin of the American Meteorological Society* 78:847-849.

Anshuka, A.; Van Ogtrop, F. and Vervoort, R. 2019. Drought forecasting through statistical models using standardised precipitation index: a systematic review and meta-regression analysis. *Natural Hazards* 97(2):955-977.

Arikan, B.; Kahya, B. and Ercan, G. 2018. Homogeneity revisited: analysis of updated precipitation series in Turkey. *Theoretical and Applied Climatology* 135(1-2):211-220. DOI: <https://doi.org/10.1007/s00704-018-2368-x>.

Arriaga, G.; Paredes, J.; Cohen, I.; Valle, M.; López, F. y Hurtado, P. 2019. Análisis temporal de sequías (1922-2016) en la cuenca alta del río Nazas usando el SPI y su relación con ENSO. *Tecnología y Ciencias del Agua* 10(5):126-153.

Ayantobo, O.; Li, Y.; Song, S.; Javed, T. and Yao, N. 2018. Probabilistic modelling of drought events in China via 2-dimensional joint copula. *Journal of Hydrology* 559:373-391.

Basu, R.; Singh, C. and Eslamian, S. 2017. Cause and Occurrence of Drought. s.l., CRC Press. p. 137-148.

Beguería, S. and Vicente-Serrano, S. 2013. SPEI: calculation of the standardised precipitation-evapotranspiration index. R package version 1:6.

Bethea, R. 1995. Statistical methods for engineers and scientists. 3 ed. Press, 1 ed. New York, Routledge. 684 p.

Box G. and Jenkins G. 1976. Time series analysis: forecasting and control. 1 ed. New York, Holden-Day, (Holden-Day series in time series analysis and digital processing). 575 p.

Box, G.; Jenkins, G.; Reinsel, G. and Ljung, GM. 2015. Time series analysis : forecasting and control. 5 ed. Wiley. New Jersey. 712 p.

Brockwell, P. and Davis, R. 2002. Introduction to time series and forecasting. 3 ed. 425 p. DOI: <https://doi.org/10.1007/b97391>.

Buishand, T. 1982. Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology* 58(1-2):11-27. DOI: [https://doi.org/10.1016/0022-1694\(82\)90066-X](https://doi.org/10.1016/0022-1694(82)90066-X).

Burnham, K. and Anderson, D. 2004. Multimodel inference; understanding AIC and BIC in model selection. *Sociological Methods & Research* 33(2):261-304. DOI: <https://doi.org/10.1177/0049124104268644>.

Chang, T. and Kleopa, X. 1991. A proposed method for drought monitoring. *Journal of the American Water Resources Association* 27(2):275-281. DOI: <https://doi.org/10.1111/j.1752-1688.1991.tb03132.x>.

Chen, L. and Guo, S. 2013. Drought analysis using copulas. *copulas and its application in hydrology and water resources* :97-116. DOI: https://doi.org/10.1007/978-981-13-0574-0_5.

Chen, L. and Guo, S. 2019. *Copulas and its application in hydrology and water resources*. 1ra ed. New York, Water Springer. 290 p.

Chen, L. and Guo, S. 2019. Uncertainty analysis of hydrologic forecasts based on copulas. *In Springer (ed.)*. New York, Springer, Singapore. p. 165-210 DOI: https://doi.org/10.1007/978-981-13-0574-0_8.

Chen, L.; Singh, V.; Guo, S. and Zhou, J. 2015. Copula-based method for multisite monthly and daily streamflow simulation. 528:369-384. DOI: https://doi.org/10.1007/978-981-13-0574-0_7.

Chow, V.; Maidment, D. and Mays, L. 1988. *Applied hydrology*. EEUU, McGraw-Hill. 572 p.

Dalezios, N.; Loukas, A.; Vasiliades, L. and Liakopoulos, E. 2000. Severity-duration-frequency analysis of droughts and wet periods in Greece. *Hydrological Sciences Journal* 45(5):751-769. DOI: <https://doi.org/10.1080/02626660009492375>.

Dracup, J.; Lee, K. and Paulson, E. 1980. On the statistical characteristics of drought events. *Water Resources Research* 16(2):289-296. DOI: <https://doi.org/10.1029/WR016i002p00289>.

Edwards, D. 1997. Characteristics of 20th century drought in the United States at multiple time scales. s.l., Air Force Inst Of Tech Wright-Patterson AFB-OH. 174 p.

Gálvez, A. 2017. Caracterización de las sequías históricas y proyectadas bajo escenarios de cambio climático en la cuenca del río Mantaro. s.l., Universidad Nacional Agraria la Molina. 106 p.

Grami, A. 2017. Probability, random variables, statistics, and random processes : fundamentals and applications. USA, Wiley. 416 p.

Hipel, K. and McLeod, A. 1994. Time series modelling of water resources and environmental systems. 1 ed. New York, Elsevier. 1013 p.

Karamouz, M.; Nazif, S.; Falahi, M.; Nazif, S. and Falahi, M. 2012. Hydrology and hydroclimatology. 1ra ed. Colombia, CRC Press. 740 p. DOI: <https://doi.org/10.1201/b13771>.

Kendall, M. 1975. Rank correlation methods. San Francisco, CA 8.

Kim, T.; Valdés, J. and Yoo, C. 2006. Nonparametric approach for bivariate drought characterization using Palmer drought index. Journal of Hydrologic Engineering 11(2):134-143.

Kisiel, C. 1969. Time series analysis of hydrologic data. Advances in hydroscience 5:1-119. DOI: <https://doi.org/10.1016/B978-1-4831-9936-8.50007-6>.

Loaiciga, H. and Leipnik, R. 1996. Stochastic renewal model of low-flow streamflow sequences. Stochastic Hydrology and Hydraulics 10(1):65-85. DOI: <https://doi.org/10.1007/BF01581794>.

Machiwal, D. and Jha, M. 2012. Hydrologic time series analysis: theory and practice. s.l., Springer Science & Business Media.

Maity, R. 2018. Statistical methods in hydrology and hydroclimatology. 1er ed. Springer Singapore. India, Springer Singapore. 444 p. DOI: <https://doi.org/978-981-10-8779-0>.

Mann, H. 1945. Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society* :245-259.

Mathier, L.; Perreault, L.; Bobée, B. and Ashkar, F. 1992. The use of geometric and gamma-related distributions for frequency analysis of water deficit. *Stochastic Hydrology and Hydraulics* 6(4):239-254.

MATLAB (Matrix Laboratory). 2020. Inc. MATLAB. Version 2020a. U.S., 1er. Disponible en www.mathworks.com/.

McKee, T.; Doesken, N. and Kleist, J. 1993. The relationship of drought frequency and duration to time scales. *In Proceedings of the 8th Conference on Applied Climatology*. s.l., American Meteorological Society Boston, MA. p. 179-183.

MINAG (Ministerio de Agricultura). (2010). Evaluación de recursos hídricos superficiales en la cuenca del río Mantaro. Perú. Lima Perú, s.e.

Mirabbasi, R.; Fakheri-Fard, A. and Dinpashoh, Y. 2012. Bivariate drought frequency analysis using the copula method. *Theoretical and Applied Climatology* 108(1-2):191-206. DOI: <https://doi.org/10.1007/s00704-011-0524-7>.

Mishra, A. and Desai, V. 2005. Drought forecasting using stochastic models. *Stochastic Environmental Research and Risk Assessment* 19(5):326-339.

Mishra, A. and Singh, V. 2010. A review of drought concepts. *Journal of Hydrology* 391(1-2):202-216. DOI: <https://doi.org/10.1016/J.JHYDROL.2010.07.012>.

Montaseri, M.; Amirataee, B. and Rezaie, H. 2018. New approach in bivariate drought duration and severity analysis. *Journal of hydrology* 559:166-181.

Organización Meteorológica Mundial. 2006. Vigilancia y alerta temprana de la sequía. Suiza, OMM, vol.OMM-N° 100. p. 28.

Palit, A. and Popović, D. 2005. *Computational intelligence in time series forecasting : theory and engineering applications*. 1st ed. London, Springer-Verlag. 372 p.

Palmer, W. 1965. *Meteorological drought*, Research paper no. 45. US Weather Bureau, Washington, DC 58.

Pankratz, A. 2009. *Forecasting with univariate Box-Jenkins models : concepts and cases*. 1st ed. New York, John Wiley & Sons. 588 p.

Panu, U. and Sharma, TC. 2009. Analysis of annual hydrological droughts: the case of northwest Ontario, Canada. *Hydrological Sciences Journal* 54(1):29-42. DOI: <https://doi.org/10.1623/hysj.54.1.29>.

Pettitt, A. 1979. A non-parametric approach to the change-point problem (en línea). *Applied Statistics* 28(2):126. DOI: <https://doi.org/10.2307/2346729>.

Pokojski, W. and Pokojska, P. 2018. Voronoi diagrams–inventor, method, applications. *Polish Cartographical Review* 50(3):141-150.

RAE (Real Academia Española). 1999. *Ortografía de la lengua española* (en línea). 1ra ed. España, RRA. 165 p. Consultado 20 feb. 2019. Disponible en <http://atlas.umss.edu.bo:8080/jspui/handle/123456789/742>.

Salas, J.; Yevjevich, V. and Lane, W. 1980. Applied modeling of hydrologic time series. 1st ed. U. S. A., Water Resources Publications. 484 p.

Sen, Z. 1976. Wet and dry periods of annual flow series (en línea). Journal of the Hydraulics Division 102(10):1503-1514. Disponible en <https://cedb.asce.org/CEDBsearch/record.jsp?dockey=0006955>.

Şenaut, Z. 2015. Applied drought modeling, prediction, and mitigation. 1st ed. Amsterdam, Elsevier. 472 p.

Sharma, TC. 1997. A drought frequency formula (en línea). Hydrological Sciences Journal 42(6):803-814. DOI: <https://doi.org/10.1080/02626669709492080>.

Shiau, J.; Feng, S. and Nadarajah, S. 2007. Assessment of hydrological droughts for the Yellow River, China, using copulas. Hydrological Processes 21(16):2157-2163. DOI: <https://doi.org/10.1002/hyp.6400>.

Shiau, J. and Shen, H. 2001. Recurrence analysis of hydrologic droughts of differing severity. Journal of Water Resources Planning and Management 127(1):30-40. DOI: [https://doi.org/10.1061/\(ASCE\)0733-9496\(2001\)127:1\(30\)](https://doi.org/10.1061/(ASCE)0733-9496(2001)127:1(30)).

Shiau, J. 2006. Fitting drought duration and severity with two-dimensional copulas. water resources management 20(5):795-815. DOI: <https://doi.org/10.1007/s11269-005-9008-9>.

Shiau, J. and Modarres, R. 2009. Copula-based drought severity-duration-frequency analysis in Iran. Meteorological Applications 16(4):481-489. DOI: <https://doi.org/10.1002/met.145>.

Shiau, J. and Shen, H. 2001. Recurrence analysis of hydrologic droughts of differing severity. *Journal of Water Resources Planning and Management* 127(1):30-40. DOI: [https://doi.org/10.1061/\(ASCE\)0733-9496\(2001\)127:1\(30\)](https://doi.org/10.1061/(ASCE)0733-9496(2001)127:1(30)).

Sirdaş, S. and Sen, Z. 2003. Spatio-temporal drought analysis in the Trakya region, Turkey. *Hydrological Sciences Journal* 48(5):809-820. DOI: <https://doi.org/10.1623/hysj.48.5.809.51458>.

Sklar, M. 1959. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris* 8:229-231.

Svoboda, M.; Hayes, M. and Wood, D. 2012. Índice normalizado de precipitación. guía del usuario. Organización Meteorológica Mundial :1-23.

Tirivarombo, S.; Osupile, D. and Eliasson, P. 2018. Drought monitoring and analysis: standardised precipitation evapotranspiration index (SPEI) and standardised precipitation index (SPI). *Physics and Chemistry of the Earth, Parts A/B/C* 106:1-10.

Tosunoğlu, F. and Onof, C. 2017. Joint modelling of drought characteristics derived from historical and synthetic rainfalls: Application of Generalized Linear Models and Copulas. *Journal of Hydrology: Regional Studies* 14:167-181. DOI: <https://doi.org/10.1016/J.EJRH.2017.11.001>.

WMO (World Meteorological Organization). (2005). informe sobre el estado del clima mundial

Wijngaard, J.; Klein Tank, A. and Können, G. 2003. Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology* 23(6):679-692. DOI: <https://doi.org/10.1002/joc.906>.

Wilhite, D. and Glantz, M. 1985. Understanding: the drought phenomenon: The role of definitions. *Water International* 10(3):111-120. DOI: <https://doi.org/10.1080/02508068508686328>.

Yevjevich, V. 1967. Objective approach to definitions and investigations of continental hydrologic droughts. *Hydrology Paper 23*, Colorado State U, Fort Collins, Aug 1967. 19 p, 9 fig, 1 tab, 12 ref. .

Yolanda, C. y Alfonso, G. 2015. *Análisis estadístico de datos espaciales con QGIS y R*. Madrid - España, Editorial UNED. 183 p.

Yoo, J.; Kim, U. and Kim, T. 2013. Bivariate drought frequency curves and confidence intervals: A case study using monthly rainfall generation. *Stochastic Environmental Research and Risk Assessment* 27(1):285-295. DOI: <https://doi.org/10.1007/s00477-012-0588-7>.

Yue, S.; Pilon, P. and Cavadias, G. 2002. Power of the Mann–Kendall and Spearman’s rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology* 259(1-4):254-271. DOI: [https://doi.org/10.1016/S0022-1694\(01\)00594-7](https://doi.org/10.1016/S0022-1694(01)00594-7).

Zargar, A; Sadiq, R; Naser, B; Khan, FI. 2011. A review of drought indices. *Environmental Reviews* 19(1):333-349. DOI: <https://doi.org/10.1139/a11-013>.

Zubieta, R.; Saavedra, M.; Silva, and; Giráldez, L. 2017. Spatial analysis and temporal trends of daily precipitation concentration in the Mantaro River basin: central Andes of Peru. *Stochastic Environmental Research and Risk Assessment* 31(6):1305-1318. DOI: <https://doi.org/10.1007/s00477-016-1235-5>.

ANEXOS

7.1 Estaciones meteorológicas.

N°	Nombres	N° Datos	Prueba de homogeneidad				Prueba de Tendencias estacional
			V – N	Bh	H-N	Pt	
1	Acobamba		1.037	0.41	0.078	1.08	0.678
2	Atocsaico		0.903	0.087	0.298	0.178	0.126
3	C.H. malpaso	732	0.903	0.882	0.027	0.712	0.073
4	Carhuamayo	600	0.817	0.516	0.498	0.447	0.660
5	Casapato	456	0.932	0.425	0.069	0.488	0.020
6	Cercapuquio	288	0.603	0.029	0.009	0.116	0.036
7	Chilicocha	204	0.794	0.169	0.119	0.074	7.854e-06
8	Choclococha	180	0.746	2.2e-16	5e-05	0.0002	2.73e-05
9	Corpacancha	228	0.561	0.963	0.739	0.811	0.693
10	Huamanga	300	0.966	0.289	0.309	0.036	0.002
11	Huancalpi	240	0.889	0.043	0.202	0.265	0.087
12	Huancavelica	180	0.999	0.726	0.141	1.046	0.942
13	Huayao	444	0.844	0.532	0.921	0.738	0.141
14	Hueghue	444	0.942	0.702	0.722	0.281	0.016
15	Huichicocha	180	0.709	0.409	0.0005	0.135	0.015
16	Ingenio	168	0.912	0.634	0.781	0.885	0.516
17	Jauja	228	0.822	0.999	0.36	1.550	0.550
18	La Oroya	228	0.828	0.999	0.768	1.416	0.563
19	La Quinua	264	0.761	0.657	0.311	1.116	0.194
20	Lircay	228	0.774	0.426	0.486	1.098	0.769
21	Luricocha	180	0.809	0.442	0.056	0.398	0.153
22	Marcapomacocha	204	0.854	0.391	0.658	0.474	0.005
23	Matibamba	168	0.943	0.815	0.838	1.283	0.899
24	Morococha	276	1.000	0.803	0.138	1.162	0.030
25	Pachacayo	444	0.784	0.505	0.657	0.588	0.077
26	Pachachaca	204	0.761	0.970	0.923	1.206	0.173
27	Palaco	288	0.931	0.678	0.798	1.203	0.504
28	Pampas	252	1.037	0.997	0.702	1.252	0.972
29	Paragsha	204	0.643	0.738	0.653	0.782	0.063
30	Paucarbamba	204	1.143	0.427	0.069	0.697	0.359
31	Pilchaca	276	0.810	0.801	0.182	1.245	0.481
32	Presa Upamayo	600	0.807	0.060	0.221	0.249	0.015
33	Salcabamba	264	0.850	2.2e-16	0.0078	0.0119	1.249e-05
34	San Pedro de Cachi	240	0.829	2.2e-16	2.2e-16	1.824e	1.524e-09
35	Shelby	252	0.918	0.708	0.847	1.291	0.733
36	Tambo del Sol	396	0.854	0.006	0.051	0.086	0.016
37	Telleria	252	0.913	0.535	0.024	0.791	0.451
38	Upamayo	252	0.913	0.544	0.025	0.791	0.458
39	Viques	264	0.761	0.211	0.161	0.533	0.052
40	Wallyapampa	204	1.004	0.152	0.302	0.813	0.857
41	Yatac	264	0.834	0.492	0.496	0.287	1.443e-05
42	Yauli	204	0.761	0.9735	0.925	1.206	0.173
43	Yauricocha	288	0.603	0.031	0.008	0.116	0.036

Nota, resultado del análisis de homogeneidad y tendencias, se muestra los p valores para las pruebas estadísticas.

7.2 Gráficos de series de tiempo

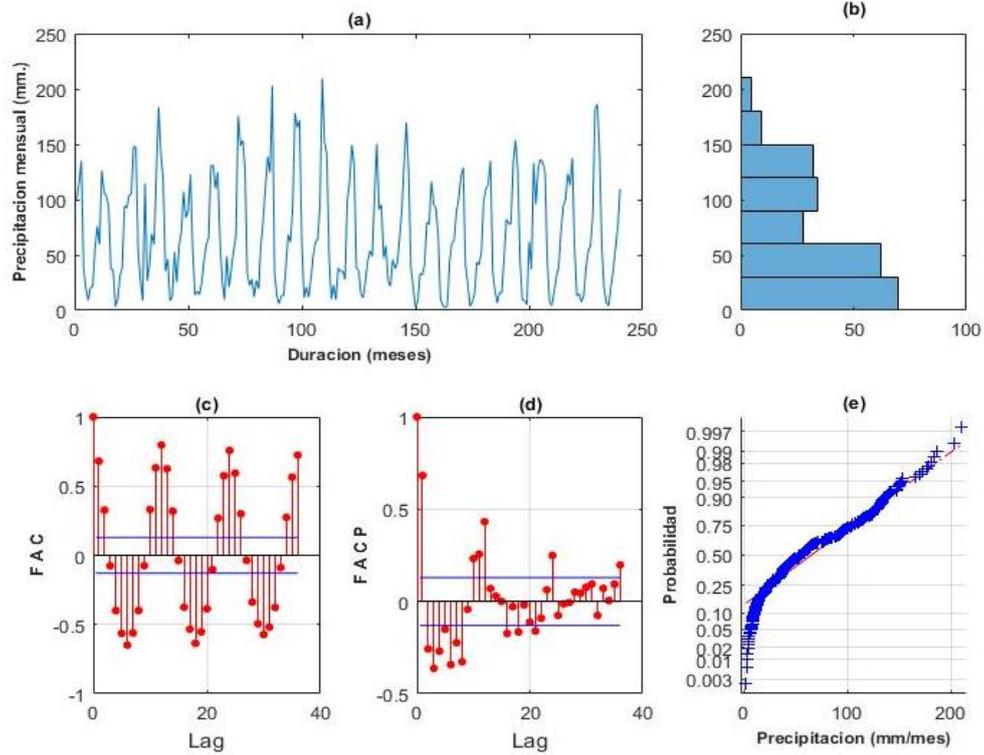


Figura 21: Serie de precipitación, zona homogénea Cuenca Baja.

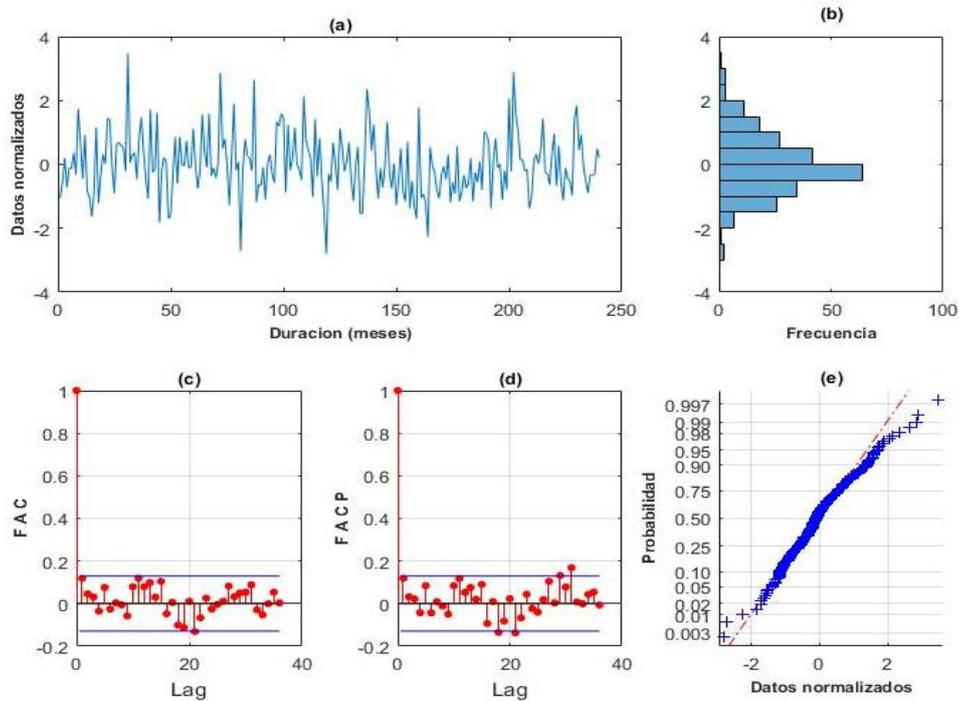


Figura 22: Serie estandarizada, zona homogénea Cuenca Baja.

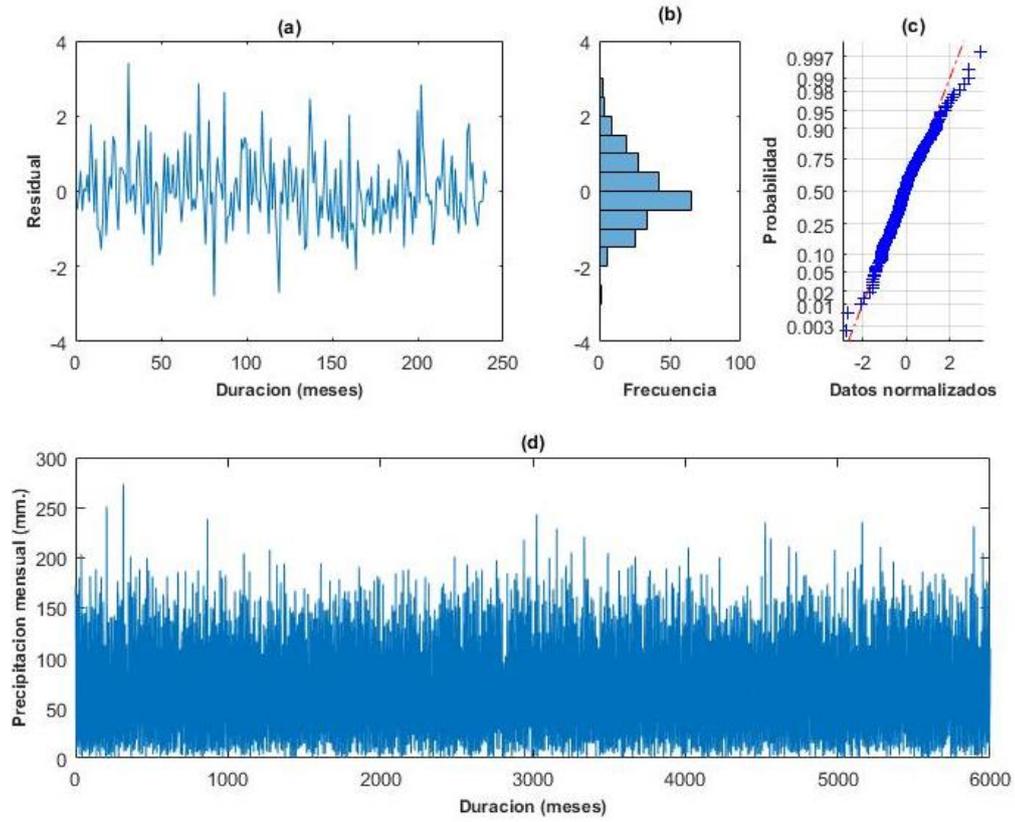


Figura 23: Serie generada, zona homogénea Cuenca Baja.

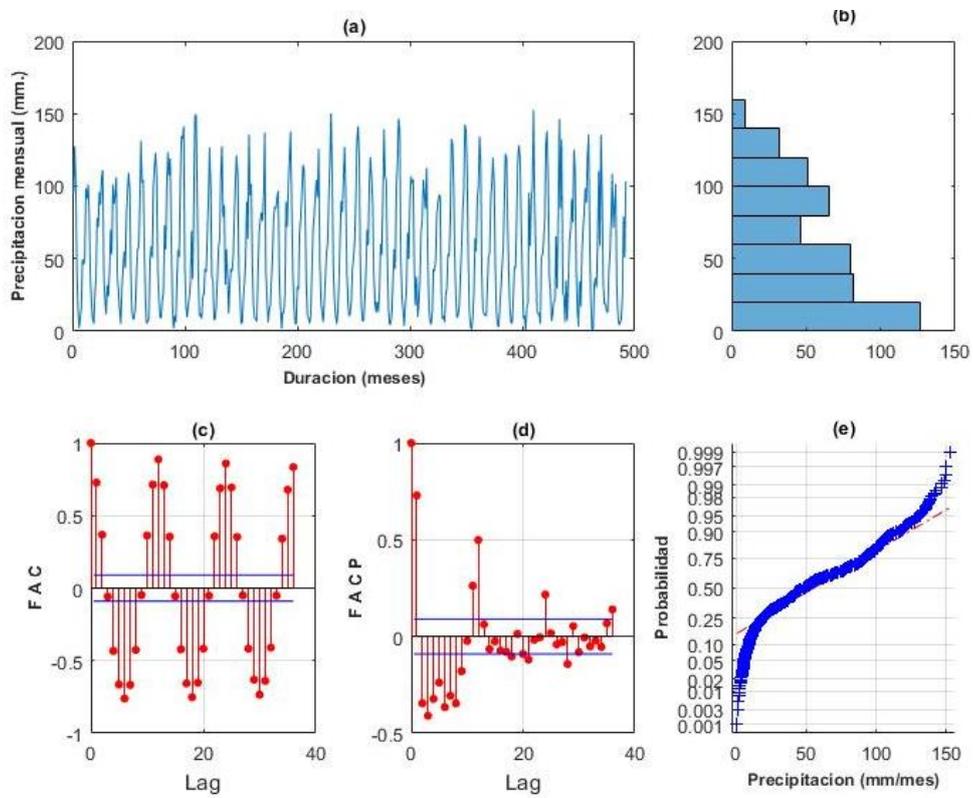


Figura 24: Serie de precipitación, zona homogénea Cuenca Media Este.

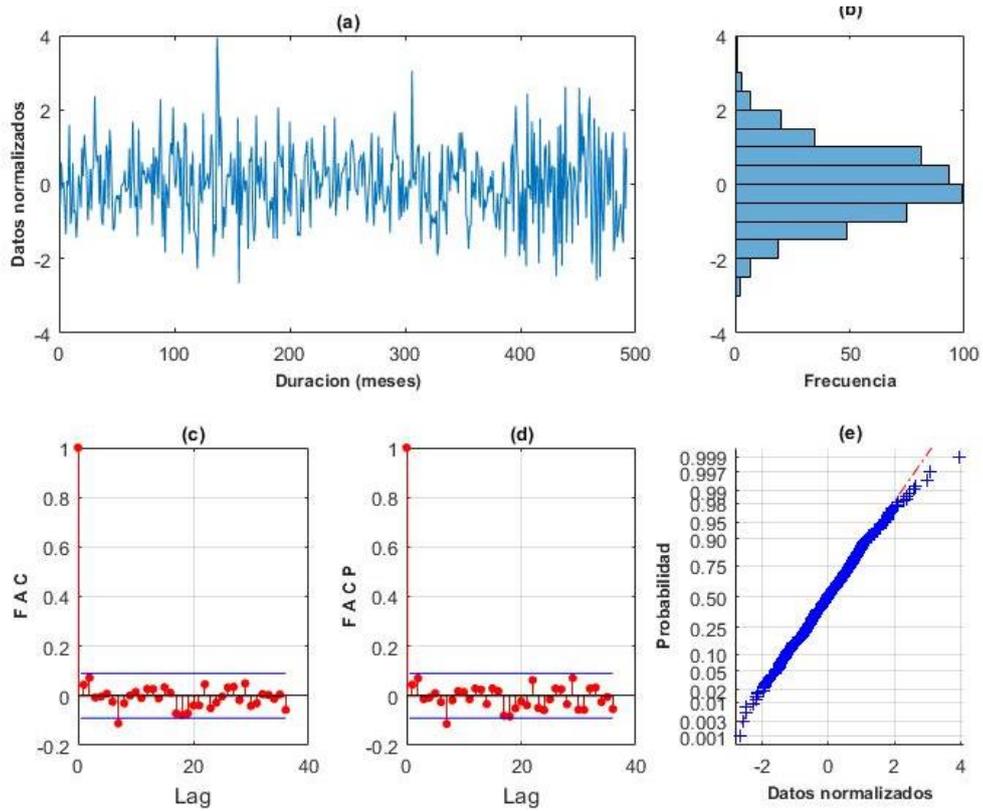


Figura 25: Serie estandarizada, zona homogénea Cuenca Media Este.

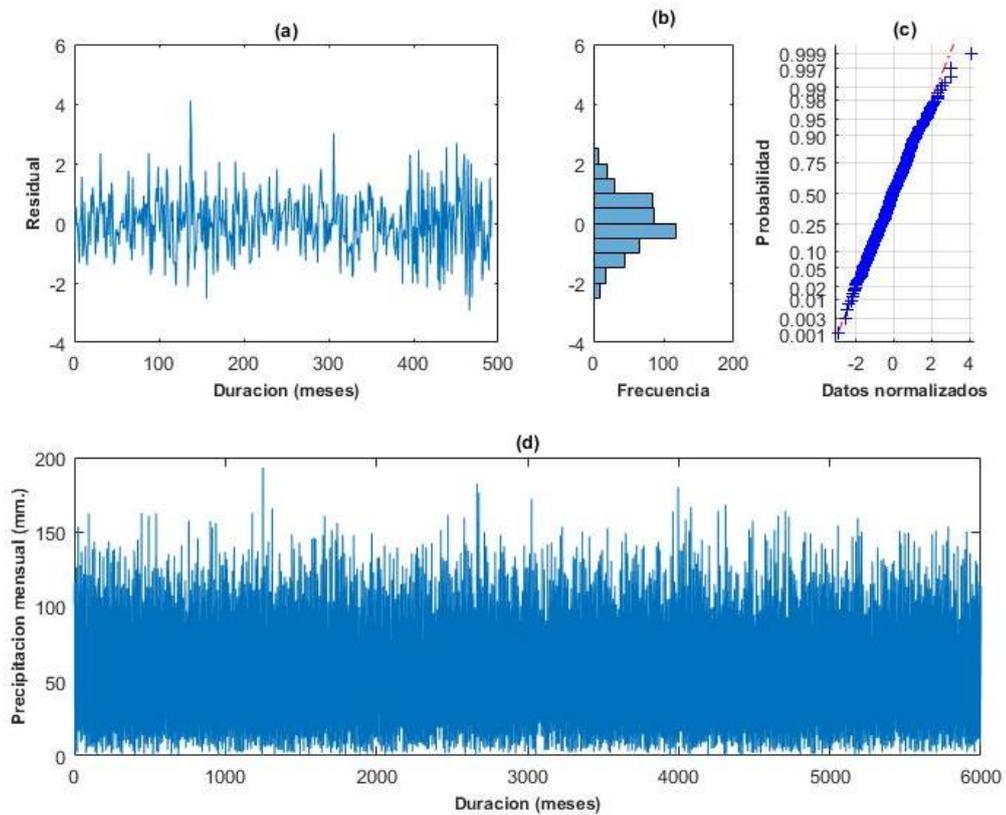


Figura 26: Serie generada, zona homogénea Cuenca Media Este.

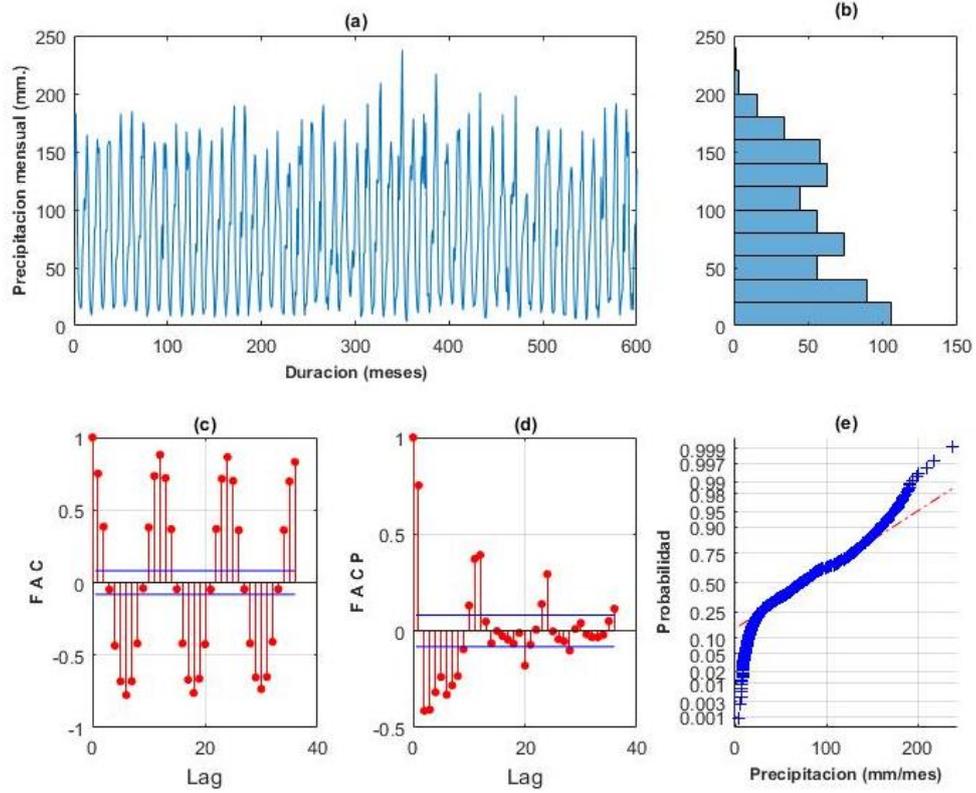


Figura 27: Serie de precipitación, zona homogénea Cuenca Alta.

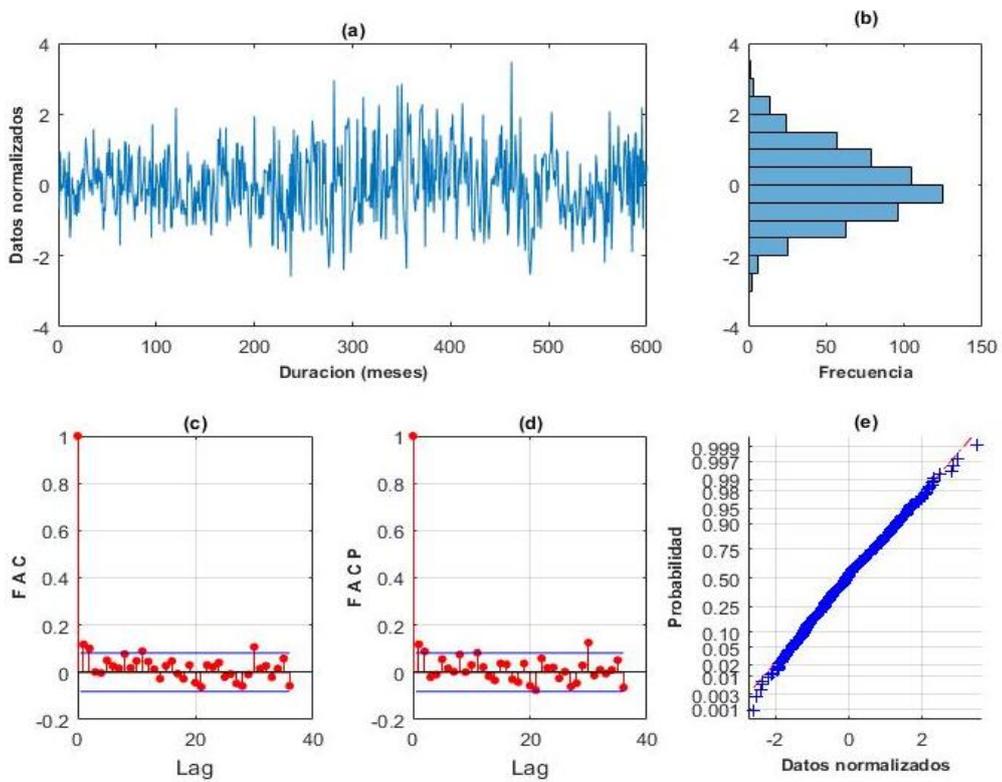


Figura 28: Serie estandarizada, zona homogénea Cuenca Alta.

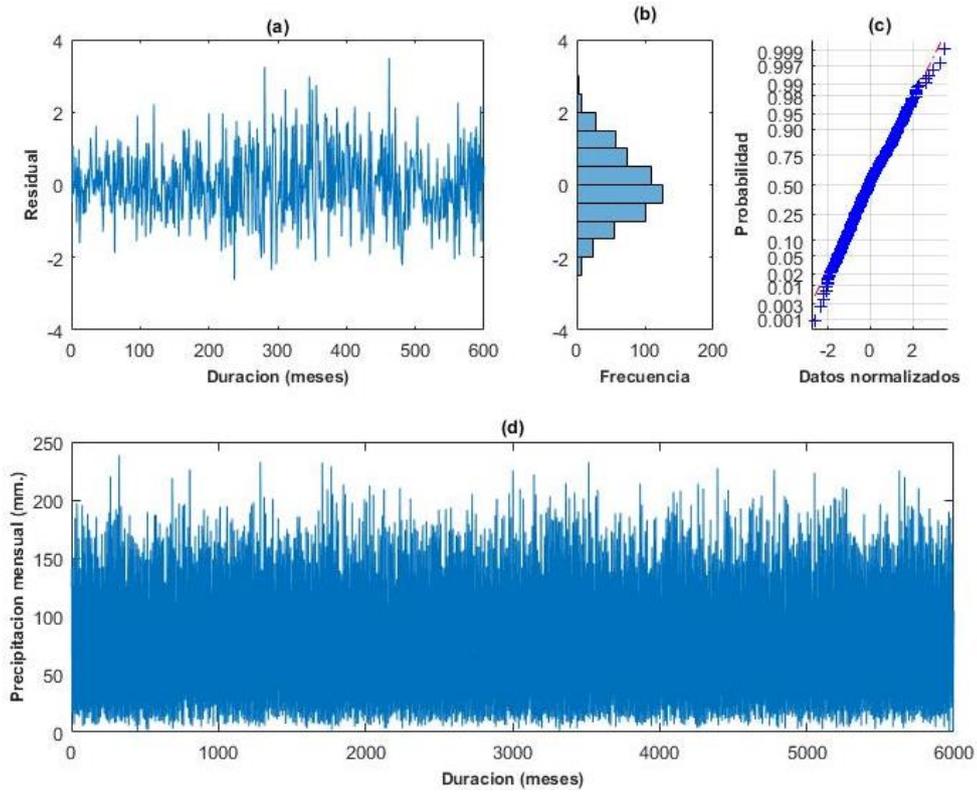


Figura 29: Serie generada, zona homogénea Cuenca Alta.

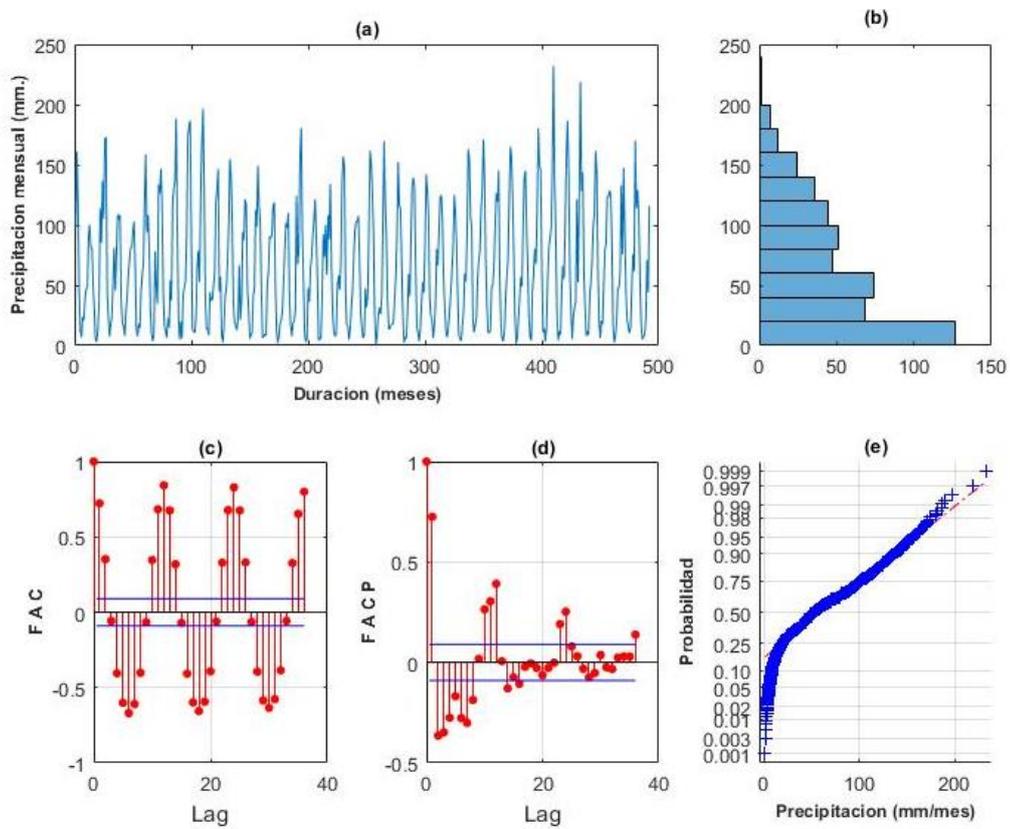


Figura 30: Serie de precipitación, zona homogénea Cuenca Media Oeste.

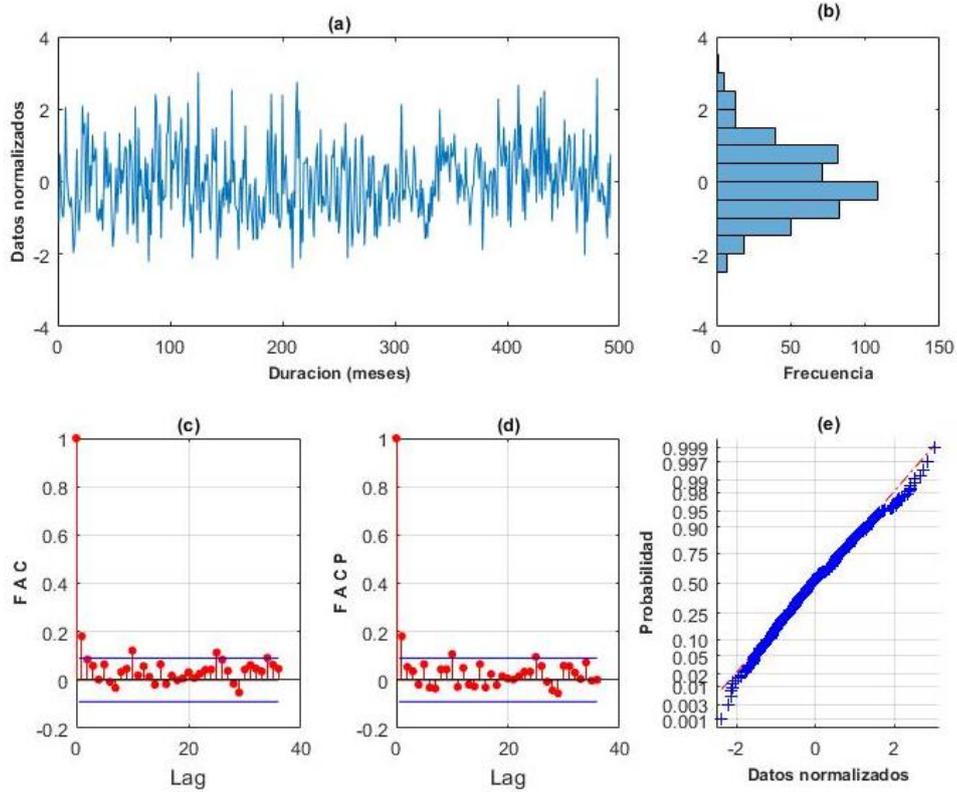


Figura 31: Serie estandarizada, zona homogénea Cuenca Media Oeste

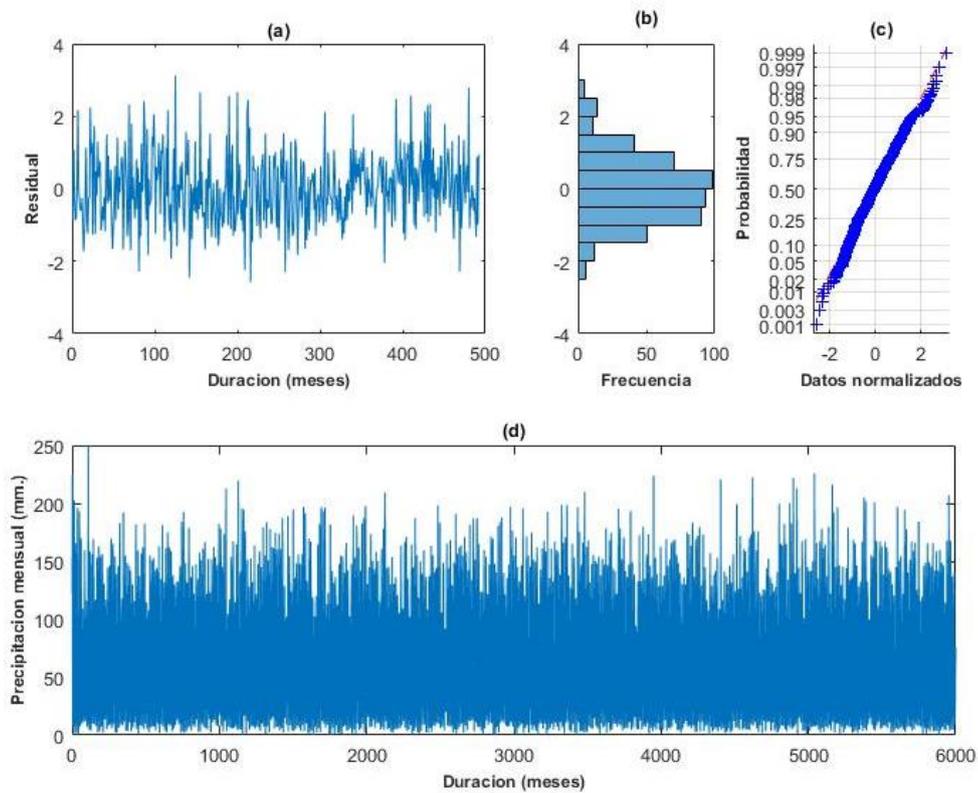


Figura 32: Serie generada, zona homogénea Cuenca Media Oeste

7.3 Ajuste de curvas, para las cuatro zonas.

Tabla 12: Parámetros de las FDP, para la zona 1; Cuenca Baja.

Modelo	p -value Parámetros	SPI (3)				SPI (12)			
		Tes	DUR	SEV	INT	Tes	DUR	SEV	INT
Gamma	RMSE	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.00
	β	1.22	2.68	1.90	10.90	0.74	1.13	0.87	10.68
	α	9.25	0.79	1.72	0.15	37.78	4.70	9.43	0.14
Log normal	RMSE	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.00
	β	1.96	0.55	0.90	0.45	2.53	1.17	1.44	0.39
	α	1.05	0.60	0.72	0.30	1.50	1.02	1.17	0.30
Gumbel	RMSE	0.02	0.02	0.02	0.01	0.03	0.02	0.02	0.01
	β	17.16	2.99	4.98	1.93	50.62	8.36	14.13	1.83
	α	15.05	2.21	4.53	0.73	70.11	7.15	14.58	0.66
Pearson	RMSE	0.00	0.05	0.00	0.00	0.01	0.02	0.01	0.00
	β	0.45	4.53	1.21	0.28	0.99	4.44	2.07	0.51
	α	5.17	0.16	0.70	0.32	10.62	0.29	1.25	0.25
Weibull	λ	5.52	1.04	1.49	1.35	7.97	1.07	1.59	1.24
	RMSE	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.01
	β	11.78	2.38	3.59	1.82	25.24	5.40	7.65	1.72
Exponencial	α	1.11	1.56	1.31	3.07	0.82	1.03	0.88	3.02
	RMSE	0.00	0.02	0.01	0.05	0.00	0.01	0.01	0.05
	β	11.30	2.11	3.27	1.64	28.04	5.32	8.23	1.54
Pareto	RMSE	0.00	0.02	0.01	0.03	0.00	0.01	0.00	0.03
	β	-0.08	-0.21	-0.08	-0.48	0.18	0.02	0.32	-0.68
	α	12.19	2.52	3.53	2.16	22.88	5.20	5.76	2.33

Nota: RMSE es la raíz del error cuadrático medio, β es el parámetro de localización, α es el parámetro de escala y λ es el parámetro de forma.

Tabla 13: Parámetros de las FDP, para la zona 1; Cuenca Media Este.

Modelo	p -value Parámetros	SPI (3)				SPI (12)			
		Tes	DUR	SEV	INT	Tes	DUR	SEV	INT
Gamma	RMSE	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.01
	β	1.29	2.92	2.10	11.79	0.74	1.14	0.89	11.28
	α	8.33	0.71	1.51	0.14	31.51	4.07	7.94	0.14
Log normal	RMSE	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.00
	β	1.94	0.55	0.89	0.44	2.33	1.03	1.30	0.38
	α	1.02	0.58	0.70	0.29	1.52	0.99	1.12	0.29
Gumbel	RMSE	0.02	0.02	0.02	0.01	0.01	0.02	0.03	0.01
	β	16.22	2.83	4.57	1.89	36.82	7.41	12.31	1.80
	α	14.63	1.96	3.78	0.62	32.07	6.71	13.64	0.61
Pearson	RMSE	0.00	0.05	0.00	0.00	0.03	0.07	0.00	0.00
	β	0.38	4.94	1.16	0.19	5.12	3.02	1.91	0.54
	α	5.07	0.23	0.73	0.33	9.86	0.00	1.03	0.23
Weibull	λ	5.57	1.05	1.50	1.36	2.93	1.00	1.52	1.23
	RMSE	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.01
	β	11.33	2.34	3.51	1.80	20.99	4.67	6.55	1.70
Exponencial	α	1.15	1.67	1.41	3.29	0.82	1.02	0.88	3.13
	RMSE	0.00	0.02	0.01	0.05	0.01	0.01	0.01	0.05
	β	10.76	2.07	3.16	1.62	23.17	4.63	7.07	1.52
Pareto	RMSE	0.00	0.02	0.01	0.03	0.00	0.01	0.00	0.03
	β	-0.09	-0.23	-0.15	-0.68	0.11	0.09	0.36	-0.68
	α	11.70	2.47	3.60	2.39	20.68	4.23	4.73	2.27

Tabla 14: Parámetros de las FDP, para la zona 1; Cuenca Alta.

Modelo	ρ -value Parámetros	SPI (3)				SPI (12)			
		Tes	DUR	SEV	INT	Tes	DUR	SEV	INT
Gamma	RMSE	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00
	β	1.25	3.04	2.16	11.78	0.71	1.16	0.91	11.70
	α	9.19	0.69	1.50	0.14	37.49	4.13	7.86	0.13
Log normal	RMSE	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00
	β	1.99	0.57	0.92	0.45	2.42	1.07	1.32	0.36
	α	1.04	0.57	0.69	0.29	1.49	1.00	1.14	0.29
Gumbel	RMSE	0.02	0.02	0.02	0.01	0.03	0.02	0.02	0.01
	β	17.36	2.86	4.70	1.92	46.54	7.44	11.82	1.77
	α	14.46	1.96	3.90	0.67	53.40	6.07	10.96	0.62
Pearson	RMSE	0.00	0.04	0.00	0.00	0.00	0.03	0.01	0.00
	β	0.42	4.09	1.06	0.22	1.32	3.90	2.28	0.51
	α	5.27	0.02	0.80	0.32	7.95	0.16	0.88	0.24
Weibull	λ	5.72	1.00	1.58	1.37	5.66	1.04	1.38	1.22
	RMSE	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.01
	β	11.99	2.37	3.59	1.82	22.95	4.86	6.74	1.67
Exponencial	α	1.13	1.70	1.42	3.21	0.79	1.04	0.90	3.18
	RMSE	0.00	0.02	0.01	0.05	0.01	0.01	0.01	0.05
	β	11.46	2.10	3.23	1.63	26.44	4.77	7.14	1.50
Pareto	RMSE	0.00	0.02	0.01	0.03	0.00	0.01	0.00	0.03
	β	-0.09	-0.23	-0.15	-0.64	0.35	0.01	0.28	-0.58
	α	12.53	2.51	3.68	2.38	17.85	4.73	5.26	2.10

Nota: RMSE es la raíz del error cuadrático medio, β es el parámetro de localización, α es el parámetro de escala y λ es el parámetro de forma.

Tabla 15: Parámetros de las FDP, para la zona 1; Cuenca Media Oeste

Modelo	ρ -value Parámetros	SPI (3)				SPI (12)			
		Tes	DUR	SEV	INT	Tes	DUR	SEV	INT
Gamma	RMSE	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	β	1.17	2.74	1.88	9.92	0.70	1.13	0.90	11.10
	α	9.88	0.80	1.83	0.17	43.27	5.42	10.28	0.14
Log normal	RMSE	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	β	1.96	0.59	0.95	0.45	2.56	1.31	1.58	0.41
	α	1.08	0.60	0.74	0.31	1.55	1.04	1.18	0.30
Gumbel	RMSE	0.02	0.02	0.02	0.01	0.01	0.02	0.02	0.01
	β	17.89	3.04	5.10	1.96	49.22	9.72	15.53	1.84
	α	15.75	2.11	4.13	0.75	43.36	8.87	15.27	0.58
Pearson	RMSE	0.00	0.04	0.00	0.00	0.01	0.03	0.01	0.00
	β	0.49	4.73	1.24	0.34	1.68	5.28	2.24	0.34
	α	5.20	0.29	0.77	0.33	8.31	1.88	1.54	0.30
Weibull	λ	5.44	1.06	1.52	1.34	5.26	1.36	1.67	1.29
	RMSE	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	β	11.93	2.47	3.77	1.85	26.98	6.21	8.76	1.75
Exponencial	α	1.08	1.62	1.33	2.98	0.80	1.03	0.90	3.28
	RMSE	0.00	0.02	0.01	0.05	0.01	0.00	0.01	0.05
	β	11.57	2.20	3.44	1.65	30.47	6.12	9.26	1.57
Pareto	RMSE	0.00	0.01	0.01	0.03	0.00	0.00	0.00	0.03
	β	-0.04	-0.24	-0.14	-0.49	0.21	0.03	0.24	-0.72
	α	12.07	2.66	3.89	2.20	24.42	5.92	7.17	2.38

Nota: RMSE es la raíz del error cuadrático medio, β es el parámetro de localización, α es el parámetro de escala y λ es el parámetro de forma.

7.4 Periodo de retorno univariado y bivariado

Tabla 16; Periodo de retorno univariado y bivariado para la zona 1.

Zona 1; SPI3					Zona 1; SPI12						
Duración (meses)	Severidad	Td (años)	Ts (años)	TSyD (años)	TSoD (años)	Duración (meses)	Severidad	Td (años)	Ts (años)	TSyD (años)	TSoD (años)
3	4	4	4	5	3	5	10	6	9	9	6
4	7	9	13	22	7	10	18	15	20	23	14
6	11	65	50	400	30	14	27	32	41	61	25
7	14	204	118	2757	77	18	35	66	67	156	42
9	17	2683	254	75441	233	23	44	159	110	499	74
10	20	11078	510	618520	488	27	52	317	160	1316	116

Tabla 17; Periodo de retorno univariado y bivariado para la zona 2.

Zona 2; SPI3					Zona 2; SPI12						
Duración (meses)	Severidad	Td (años)	Ts (años)	TSyD (años)	TSoD (años)	Duración (meses)	Severidad	Td (años)	Ts (años)	TSyD (años)	TSoD (años)
3	4	4	4	5	3	5	10	6	9	9	6
4	7	10	14	27	8	10	19	17	23	29	14
6	10	110	41	583	31	14	29	35	50	89	27
7	13	449	106	5733	87	18	38	72	85	245	46
9	16	11349	248	329230	243	23	47	165	134	761	82
10	19	69482	532	4287400	528	27	56	306	198	1946	128

Tabla 18; Periodo de retorno univariado y bivariado para la zona 3.

Zona 3; SPI3					Zona 4; SPI12						
Duración (meses)	Severidad	Td (años)	Ts (años)	TSyD (años)	TSoD (años)	Duración (meses)	Severidad	Td (años)	Ts (años)	TSyD (años)	TSoD (años)
3	4	4	4	5	3	5	8	6	8	8	6
4	7	11	14	20	9	8	14	12	16	17	12
6	9	122	29	264	26	12	21	27	32	39	23
7	12	521	80	2485	72	16	28	62	57	105	41
9	15	15000	196	160440	193	19	35	115	93	213	68
10	18	99892	434	2309100	432	23	41	259	135	986	97

Tabla 19; Periodo de retorno univariado y bivariado para la zona 4.

Zona 4; SPI3					Zona 4; SPI12						
Duración (meses)	Severidad	Td (años)	Ts (años)	TSyD (años)	TSoD (años)	Duración (meses)	Severidad	Td (años)	Ts (años)	TSyD (años)	TSoD (años)
3	4	4	3	4	3	7	10	8	8	9	8
4	7	8	11	14	7	12	19	18	20	24	16
6	10	63	29	137	23	18	27	46	37	69	30
7	12	206	52	661	44	24	36	114	69	226	53
9	15	3026	112	18552	109	29	45	239	119	668	90
10	18	13479	226	158600	222	35	54	564	192	2259	153