

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

**ESCUELA DE POSGRADO
MAESTRÍA EN RECURSOS HÍDRICOS**



**“SEVERIDAD Y RIESGO DE SEQUÍA DEL RÍO CHANCAY,
LAMBAYEQUE, PERÚ”**

Presentada por:

ERLE OTTO JAVIER BUSTAMANTE SCAGLIONI

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

Lima - Perú

2022

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

**ESCUELA DE POSGRADO
MAESTRÍA EN RECURSOS HÍDRICOS**

**“SEVERIDAD Y RIESGO DE SEQUÍA DEL RÍO CHANCAY,
LAMBAYEQUE, PERÚ”**

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

Presentada por:

ERLE OTTO JAVIER BUSTAMANTE SCAGLIONI

Sustentado y aprobado ante el siguiente jurado:

Ph.D. Eduardo Chávarri Velarde
PRESIDENTE

Mg.Sc. Miguel Sánchez Delgado
ASESOR

Dr. Abel Mejía Marcacuzco
MIEMBRO

Mg.Sc. Gonzalo Fano Miranda
MIEMBRO

DEDICATORIA

A Dios, Luvia y Javier, para ellos mi amor y esfuerzo.

AGRADECIMIENTOS

Al Mg.Sc. Miguel Sánchez Delgado por su permanente apoyo, colaboración y consejos para el término de la presente tesis, además a los docentes de la maestría en Recursos Hídricos de la UNALM por compartir sus conocimientos, y por último agradecer al Proyecto Especial Olmos Tinajones PEOT y el Servicio Nacional de Meteorología e Hidrología del Perú SENAMHI por el apoyo con la información brindada.

ÍNDICE GENERAL

I.	INTRODUCCIÓN	1
1.1.	Objetivos	2
II.	REVISIÓN DE LITERATURA.....	3
2.1.	ANÁLISIS Y MODELAMIENTO HIDROLÓGICO DE SERIES DE TIEMPO.....	3
2.2.	SERIES DE TIEMPO HIDROLÓGICO.....	4
2.3.	PARTICIÓN DE LA ESTRUCTURA DE TIEMPO	5
2.3.1.	Tendencias y cambios.....	5
2.3.2.	Estacionalidad y estacionariedad.....	6
2.4.	FUNCIÓN DE DISTRIBUCIÓN NORMAL Y LA VARIABLE ESTANDARIZADA	6
2.5.	PROCESO ESTOCÁSTICO.....	7
2.6.	MODELOS LINEALES AUTORREGRESIVOS O UNIVARIADOS	8
2.6.1.	Modelo Autorregresivo AR (p)	8
2.6.2.	Modelos de promedios móviles MA (q).....	9
2.6.3.	Estimación de los parámetros del modelo	10
2.7.	PRONÓSTICO.....	13
2.7.1.	Simulación estocástica.....	14
2.7.2.	Generando series sintéticas.....	15
2.8.	SEQUÍA	16
2.8.1.	Tipos de sequías	17
2.8.2.	Antecedentes de sequías en la zona de estudio	18
2.8.3.	Propiedades de la sequía.....	21
2.8.4.	Identificación de la sequía	22
2.8.5.	Análisis de los eventos de sequía	25
2.9.	FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD DE EVENTOS DE SEQUÍA.....	27
2.9.1.	Severidad y riesgo de sequía	30
2.10.	CADENAS DE MÁRKOV	32
III.	MATERIALES Y MÉTODOS	34
3.1.	UBICACIÓN DE LA ZONA DE ESTUDIO	34

3.2.	DESCRIPCIÓN GENERAL DE LA CUENCA INTEGRADA CHANCA Y-LAMBAYEQUE.....	34
3.3.	DESCRIPCIÓN GENERAL DEL SISTEMA TINAJONES.....	36
3.4.	HIDROMETRÍA DEL RÍO CHANCA Y-LAMBAYEQUE.....	38
3.5.	METODOLOGÍA	40
	3.5.1. Selección de modelo de serie de tiempo.....	40
	3.5.2. Análisis de los eventos de sequía	51
IV.	RESULTADOS Y DISCUSIÓN	55
	4.1. SELECCIÓN DE MODELO DE SERIE DE TIEMPO.....	55
	4.2. ANÁLISIS DE LOS EVENTOS DE SEQUÍA.....	64
	4.3. ESTIMACIÓN DE LA PROBABILIDAD DE EXCEDENCIA Y RIESGO DE LOS EVENTOS DE SEQUÍA	68
V.	CONCLUSIONES	77
VI.	RECOMENDACIONES	78
VII.	REFERENCIAS BIBLIOGRÁFICAS.....	79
VIII.	ANEXOS	84

ÍNDICE DE CUADROS

Cuadro 1. Características de las sequías en términos de intensidad y severidad.....	20
Cuadro 2. Eventos climáticos en el Perú 1950-1998.....	20
Cuadro 3. Volúmenes anuales registrados en la estación Raca Rumi.	39
Cuadro 4. Prueba en la media y desviación estándar primer y segundo periodo de la serie.....	56
Cuadro 5. Prueba en la media y desviación estándar segundo periodo de la serie.....	56
Cuadro 6. Resumen de la prueba en el coeficiente de correlación.	56
Cuadro 7. Estadísticos de la serie histórica y transformada de los volúmenes del río Chancay-Lambayeque (MMC) periodo 1970 -2019.....	57
Cuadro 8. Prueba de asimetría y curtosis.	58
Cuadro 9. Prueba de Filliben en la serie de histórica y transformada.	59
Cuadro 10. Prueba de significancia de la función de autocorrelación.....	61
Cuadro 11. Parámetros de los modelos univariados autorregresivos y media móviles.....	61
Cuadro 12. Ecuación de los modelos univariados.....	62
Cuadro 13. Medidas de evaluación de los modelos univariados.....	62
Cuadro 14. Prueba de Jarque Bera a las técnicas de pronóstico seleccionados.....	63
Cuadro 15. Criterio de información de Akaike (AIC) y Schwarz (SBC).....	63
Cuadro 16. Estadísticos de la serie sintética de los volúmenes anuales del río Chancay-Lambayeque (MMC).....	64
Cuadro 17. Prueba en la media y desviación estándar.	64
Cuadro 18. Prueba en el coeficiente de correlación.	64
Cuadro 19. Nivel de truncamiento de la demanda de agua.	65
Cuadro 20. Eventos de sequía ocurridos Do de un año de duración lo.	66
Cuadro 21. Eventos de sequía observados.....	67
Cuadro 22. Prueba de Grubs a los déficits generados.	67
Cuadro 23. Estimación de las probabilidades de transición de los estados.....	68
Cuadro 24. Prueba de Anderson – Darling a los déficits generados.	68
Cuadro 25. Parámetros de la función de distribución de probabilidad Gamma.....	69
Cuadro 26. Probabilidad, periodo de retorno y riesgo de los eventos de sequía.	70
Cuadro 27. Valores críticos de la prueba de Grubbs.	89
Cuadro 28. Valores críticos para la prueba Anderson-Darling a una distribución Gamma.....	91

ÍNDICE DE FIGURAS

Figura 1. Suministro de agua x_t , demanda de agua x_0 , y propiedades de la sequía: (a) duración L; (b) déficit acumulado D; (c) intensidad I.....	21
Figura 2. Nivel de truncamiento.....	24
Figura 3. Caudal máximo, caudal mínimo, sequía, inundación y el periodo integral.....	24
Figura 4. Definición esquemática de los tiempos de retorno $T_i(j), j = 1, 2, \dots$ de cuatro tipos de eventos de sequía: (1) $\{D > D_0 \text{ y } L = 3\}$; (2) $\{D > D_0 \text{ y } L \geq 3\}$; (4) $\{I \geq I_0 \text{ y } L \geq 3\}$; (3) $\{I > I_0 \text{ y } L = 3\}$	25
Figura 5. Zona de estudio – Cuenca Chancay Lambayeque.....	38
Figura 6. Esquema de la metodología para la selección del modelo de serie de tiempo y análisis de los eventos de sequía.....	40
Figura 7. Hidrograma histórico registrado en la Estación Raca Rumi, Lambayeque.....	55
Figura 8. Histograma del volumen anual MMC de la estación Raca Rumi.....	57
Figura 9. Diagrama de cuantiles de la serie histórica y transformada.....	59
Figura 10. Diagrama de caja serie histórico y transformada.....	60
Figura 11. Función de autocorrelación FAC de la serie observada y transformada.....	60
Figura 12. Comparación serie simulada vs serie observada.....	62
Figura 13. Probabilidad de los eventos de déficit de sequía MMC de la muestra generada del río Chancay Lambayeque respecto a los déficits observados Do.....	71
Figura 14. Periodo de retorno de los eventos de déficit de sequía de la muestra generada del río Chancay Lambayeque respecto a los déficits observados Do.....	71
Figura 15. Riesgo de los eventos de déficit de sequía MMC de la muestra generada del río Chancay Lambayeque respecto a los déficits observados Do.....	72

ÍNDICE DE ANEXOS

Anexo 1. Serie histórica de volúmenes anuales de descargas del río Chancay Lambayeque. Estación Raca Rumi en millones de metros cúbicos 1914 – 2019.	84
Anexo 2. Serie histórica de rachas de sequía por coeficientes de déficit $g=0.85,0.95,1.00$ y 1.15	85
Anexo 3. Estadísticos de 10 series generadas de 50 años cada una en millones de metros cúbicos MA (1)	86
Anexo 4. Estadísticos de 10 series generadas de 1500 años cada una en millones de metros cúbicos MA (1)	87
Anexo 5. Prueba estadística de Grubbs	88
Anexo 6. Prueba estadística de Anderson – Darling para una PDF Gamma	90
Anexo 7. Déficit observado en la estación Raca Rumi 1970-2019.	91
Anexo 8. Resumen de la prueba de ajuste Anderson Darling y estimación de parámetros PDF Gamma para los déficits generados.	92
Anexo 9. Conteo de estados de falla y seguro y sus probabilidades de transición por coeficientes de déficit de la serie simulada y observada.	93

RESUMEN

Los déficits de sequía de los volúmenes de agua del río Chancay Lambayeque para el periodo 1970-2019, a menudo ocurren cada año con determinada severidad y riesgo de exposición de la demanda de agua de uso poblacional y agrícola. La incertidumbre de la frecuencia y el periodo de tiempo de la falta de agua excedan está presente. El objetivo de este estudio fue estimar la severidad y riesgo de sequía en el río Chancay Lambayeque, con demandas de agua de 927.65 MMC, 1036.78 MMC, 1091.35 MMC y 1255.05 MMC. Se identificó los déficits de sequía de 1, 2, 3 y 4 años de duración basadas en rachas o corridas de las series observadas y sintéticas, previa selección de un modelo de serie de tiempo de media móviles MA (1) para el pronóstico de los volúmenes anuales a largo plazo. Finalmente, con la secuencia de años secos y húmedos como los déficits de sequía generados, se ajustaron a una distribución de probabilidad bivariado y se obtuvieron el periodo de retorno y el riesgo de las sequías. El riesgo que exceda los déficits de sequía de un año de duración son 483.6110 MMC (0.26% - 0.85%); 269.0863 MMC (0.98% - 2.3%); 206.2540 MMC (1.42% - 2.99%); 59.3485 MMC (3.37% - 5.00%); y 19.4640 MMC (4.19% - 5.45%). Los modelos probabilísticos de media móvil y de la distribución bivariado de sequía fueron válidos para estimar el déficit de sequía de un año y están sujetos a una adecuada identificación de las sequías asociadas con la periodicidad de la demanda de agua.

Palabras claves: Sequía, modelo de series temporales, probabilidad bivariado.

ABSTRACT

The drought deficits of the water volumes of the Chancay Lambayeque River for the period 1970-2019, often occur each year with a certain severity and risk of exposure to the demand for water for population and agricultural use. The uncertainty of the frequency and time period of the lack of water exceed is present. The objective of this study was to estimate the severity and risk of drought in the Chancay Lambayeque river, with water demands of 927.65 MMC, 1036.78 MMC, 1091.35 MMC and 1255.05 MMC. Drought deficits of 1, 2, 3 and 4 years duration were identified based on runs or runs of the observed and synthetic series, after selecting a time series model of moving average MA (1) for the forecast of the long-term annual volumes. Finally, with the sequence of dry and wet years as the generated drought deficits, they were adjusted to a bivariate probability distribution and the return period and the risk of droughts were obtained. The risk that exceeds the one-year drought deficits is 483.6110 MMC (0.26% - 0.85%); 269.0863 MMC (0.98% - 2.3%); 206.2540 MMC (1.42% - 2.99%); 59.3485 MMC (3.37% - 5.00%); and 19.4640 MMC (4.19% - 5.45%). The probabilistic models of the moving average and the bivariate distribution of drought were valid to estimate the drought deficit of one year and are subject to an adequate identification of the droughts associated with the periodicity of the water demand.

Keywords: Drought, time series model, bivariate probability.

I. INTRODUCCIÓN

La sequía es un fenómeno atmosférico e hidrológico complejo, evolucionan temporal y espacialmente con anomalías propias de las lluvias y el escurrimiento en una región del planeta, se expresa en escasez de agua para el desarrollo de las actividades productivas humanas sujetos a la eventualidad de su ocurrencia, inclusive en zonas inconfundiblemente lluviosas.

Los factores climáticos la hacen más grave a la sequía cuando las altas temperaturas, baja humedad juegan en el agotamiento del agua almacenada en el suelo, lagos y la determinación de la demanda del recurso hídrico, para su uso agrícola, municipal y otros tipos.

La severidad de una sequía a menudo es examinada con índices de precipitación estandarizado SPI, índice de severidad de sequía de Palmer, índice de caudal estandarizado IQE entre otros, con poca oportunidad para conocer la recurrencia de la sequía y el riesgo de eventos de sequía específicos sucedan en un horizonte de tiempo.

Los últimos años 2019, 2016 y 2014, diferentes diarios del país reportan, la región norte del país, sufre la ausencia de lluvias con sequías desde moderadas a extremas, distintas instituciones ha señalado como consecuencia de este fenómeno, la dotación de agua por parte de los operadores del sistema hidráulico Chancay – Lambayeque no es regular para el servicio poblacional, agrícola y ganadera e industrial, especialmente en la parte baja de la cuenca, aún si los registros trajeron disminución en la disponibilidad hídrica debido al retraso de las precipitaciones, para que los operadores de abastecimiento de agua reúnan suministros adecuados, deben estimar la probabilidad de excedencia y el periodo de retorno de los eventos de sequía.

A través de esta investigación se pretende responder ¿cómo pueden ser útiles los modelos probabilísticos para el análisis de la severidad y riesgo de sequía en el río Chancay Lambayeque?

1.1. Objetivos

Objetivo General

- Estimar la severidad y riesgo de sequía en el río Chancay Lambayeque.

Objetivo Específico

- Realizar el análisis de la serie de tiempo de volúmenes anuales del río Chancay Lambayeque.
- Determinar modelos estocásticos univariados que mejor representen el pronóstico de la serie de volúmenes anuales de agua.
- Estimar la probabilidad de excedencia y riesgo de los eventos de sequía.

II. REVISIÓN DE LITERATURA

2.1. ANÁLISIS Y MODELAMIENTO HIDROLÓGICO DE SERIES DE TIEMPO

El análisis de series de tiempo se ha convertido en una importante herramienta en hidrología. Su uso se ha destacado para la construcción de modelos matemáticos, generación de registros hidrológicos sintéticos, como la completación y extensión de datos ausentes relacionados con problemas de planeamiento y manejo de sistemas de recursos hidráulicos. Por ejemplo, la determinación de la capacidad de un reservorio, o la fiabilidad de este, también la evaluación de la idoneidad de una estrategia de manejo de recursos hídricos bajo varios escenarios hidrológicos potenciales (Sveinsson *et al.* 2007).

Las series de tiempo en el modelamiento hidrológico se pueden dividir en una componente determinística la cual se usa con propósitos predictivos y una componente estocástica que consiste en el efecto de la probabilidad del fenómeno.

Esta serie como un conjunto de observaciones en el que se mide la variación en el tiempo de escurrimiento cuyos valores tomados en el tiempo se llaman estados, que pueden ser variables continuas o discretas, asimismo esta variable aleatoria en un proceso estocástico manifiesta la incertidumbre del resultado de eventos cuantificados o registrados (Alvarado 2007).

La componente estocástica está constituida por oscilaciones irregulares y efectos aleatorios y para describir se requieren conceptos de probabilidad. Una característica importante que poseen las series de tiempo discretas es la estacionalidad. Para definir matemáticamente esta propiedad se denotan las funciones de probabilidad conjunta de funciones tomada en el tiempo, $t = 1, 2, \dots, N$ y nuevamente en los tiempos $t+1, t+2, \dots, t+N$, como una función $f(y_1, y_2, \dots, y_N)$ y $f(y_{t+1}, y_{t+2}, \dots, y_{t+N})$, respectivamente. Entonces para un proceso estocástico estacionario, los dos conjuntos de funciones de densidad de probabilidad de N dimensiones es idéntico para todos los *valores de N y t* .

En general son dos los modelos de series de tiempo, los paramétricos y no paramétricos. El primero requiere el cálculo de parámetros para conservar su estructura de varianza y

covarianza en el espacio y en el tiempo. El segundo modelo no requiere o requieren un mínimo de parámetros para representar las propiedades estadísticas de las series históricas, para esto, se generan funciones de distribuciones empíricas, a través de bloques de remuestreo (bootstrap) de los datos originales (Alvarado 2007).

A continuación, se resalta los conceptos y aplicaciones en que los modelos estadísticos se basan:

- Primero, si la secuencia de los registros no representa adecuadamente una serie, no se puede desarrollar un análisis confiable. Se asume también que una muestra representa una población con suficiencia, esto también implica que la estacionalidad es invocada, de la otra manera, se puede utilizar modelos estacionarios.
- Segundo, debe ser posible identificar y rectificar errores de medida y de registro, si estos existieran, en un primer análisis de los datos. También, cualquier efecto forzado en las series de tiempo hidrológicas causado por los diferentes usos del río o por la construcción de una represa se debe tomar en cuenta, ya que las series de tiempo representan los procesos naturales tal como ocurren.
- Tercero, el modelo debe ser manejable, este deberá mantener las propiedades estadísticas e hidrológicas del proceso, así que los diferentes tipos de eventos y secuencias que ocurren de igual manera se generan de forma realista.

No obstante, en la práctica estas condiciones no se pueden alcanzar completamente, y con frecuencia hay ciertas suposiciones. Con tales consideraciones en mente un modelo que sea lo suficientemente flexible y parsimonioso con respecto a sus parámetros debe satisfacer los propósitos generales para los que se construye, es decir si para un mismo fenómeno existen dos modelos que lo pueden representar, se deberá seleccionar el más sencillo, es decir, el que requiere menos parámetros para su ajuste.

2.2. SERIES DE TIEMPO HIDROLÓGICO

Los procesos hidrológicos como el escurrimiento se desarrollan en una escala de tiempo continua y proveen un registro. Por ejemplo, los registros hidrométricos de las descargas de agua de una corriente natural en una estación son el resultado de observar los valores de

los volúmenes de agua a lo largo del tiempo en intervalos regulares o escalas de tiempo discretos, diarios, semanales, mensuales o anuales (Maidment *et al.* 1993 y Peña 2010).

Por su particularidad, como el comportamiento o régimen de las aguas y el tratamiento de series de tiempo hidrológico, estas pueden ser clasificadas en varias categorías dependiendo de un número de factores.

La serie de tiempo simple o univariada es sencillamente una serie de tiempo de una variable hidrológica en un sitio dado, asimismo, esta puede mostrar correspondencia o relación en el tiempo o el precedente, luego esta serie es llamada auto correlacionada en serie o en tiempo, de otra forma es llamada serie independiente.

Además, una serie hidrológica es estacionaria si está libre de tendencia, cambios o periodicidad, es decir la serie oscila alrededor de un nivel constante esto implica que los parámetros estadísticos de la serie tales como la media y varianza permanezcan constantes a través del tiempo. Generalmente se tiene series estacionarias aquellas definidas a una escala de tiempo anual, asunción errónea como resultado de la variabilidad climática de larga escala, interrupciones naturales o cambios inducido por el hombre (Maidment *et al.* 1993).

2.3. PARTICIÓN DE LA ESTRUCTURA DE TIEMPO

Las series hidrológicas exhiben en varios grados tendencias, cambios o saltos, estacionalidad, autocorrelación y no normalidad. Estos atributos son referidos como componentes y, una serie de tiempo puede ser particionada o descompuesta en sus series componentes (Maidment *et al.* 1993).

2.3.1. Tendencias y cambios

Las tendencias y cambios en la serie de tiempo son incluidos por factores humanos o naturales instantáneos o graduales, por ejemplo, la afectación del escurrimiento luego de la pérdida de cobertura vegetal produce súbito incremento de las aguas por corto periodo, mientras una gradual destrucción del bosque puede resultar con cambios graduales o tendencias en la serie hidrológica por un largo periodo.

Estas tendencias y cambios en la serie se expresan en una o más de sus características estadísticas, frecuentemente las tendencias en la media y en la varianza. Una tendencia

lineal en la media puede ser removida por la diferencia del valor observado y la media de la variable. Mientras si tales diferencias persisten, las tendencias en la varianza pueden ser removida dividiéndola por la desviación estándar.

De manera que, de este proceso de construir una serie por sustracción de la media y/o estandarizada según corresponda el análisis de la serie en cuestión, la serie residual mostrara media cero y varianza uno que, además puede exhibir otra propiedad como la autocorrelación.

2.3.2. Estacionalidad y estacionariedad

Las series hidrológicas que no se mantienen en un nivel constante decimos que son series no estacionarias (Peña 2010); de otra forma si la serie es definida en intervalos de tiempo más pequeñas que un año, como series mensuales generalmente exhiben patrones estacionales o periódicos resultado del movimiento anual de la tierra alrededor del sol (Maidment *et al.* 1993). Estos patrones, estacionales, se traducen en características estadísticas en la media, varianza, covarianza y asimetría que varían en el año.

Por el contrario, una serie anual es estacionaria, cuando las propiedades de las series no cambian en el tiempo absoluto (tiempo invariante) es decir no muestran un patrón periódico, oscila en un nivel constante alrededor de una media constante, las características estadísticas como la media, varianza, covarianza y asimetría no varían (Aliaga 1985 y Peña 2010).

La estacionariedad es importante en series hidrológicas, principalmente por dos razones: las técnicas matemáticas para el análisis de estas series están bien desarrolladas, y porque la mayoría de las componentes estocásticas de las series hidrológicas pueden ser consideradas aproximadamente estacionarias, una vez que las componentes deterministas conocidas han sido identificadas, sus parámetros son estimados y las tendencias o cambios son removidos desde la serie (Aliaga 1985).

2.4. FUNCIÓN DE DISTRIBUCIÓN NORMAL Y LA VARIABLE ESTANDARIZADA

De acuerdo con Alvarado (2007), la función de distribución de probabilidad Normal es definida como una función de probabilidades de una variable aleatoria continua z :

$$f(z) = \frac{1}{\sigma_z \sqrt{2\pi}} e^{-(z-\mu_z)^2/2\sigma^2} \quad (1)$$

En la que μ_z y σ_z^2 son la media y la varianza de z respectivamente, comúnmente se denota como $N(\mu, \sigma)$. Los parámetros se estiman con la muestra disponible. Si se hace un cambio de variable: $t = (z - \mu)/\sigma$, la ecuación (1) queda como sigue:

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \quad (2)$$

Que se conoce como normal estándar, donde (2) representa una función sin parámetros, es decir, $\mu = 0$ y $\sigma^2 = 1$ que se denota como $N(0,1)$. Esta función se utiliza comúnmente en hidrología, particularmente en modelos de series de tiempo como función de densidad de probabilidades para el análisis de errores aleatorios. Generalmente para el caso de volúmenes medios anuales, la función de densidad de probabilidad con transformación logarítmica log-normal, se aplica para estandarizar esta variable, (Yevjevich 1972 citado por Alvarado 2007).

2.5. PROCESO ESTOCÁSTICO

En las series de tiempo los valores se ordenan en función del tiempo, y generalmente es el caso de que el valor observado de la variable aleatoria en un tiempo dado influye en la distribución de la variable aleatoria en tiempos posteriores; en tal sentido, las observaciones no son independientes.

Las propiedades de un proceso estocástico se deben determinar por una serie de tiempo simple o realización. Para hacer esto, primeramente, se hacen ciertas suposiciones. Primero se asume que el proceso es estacionario, eso significa que la función de distribución de probabilidad del fenómeno no cambia a través del tiempo. Dicho de otra manera: si X_t es un proceso estacionario, entonces:

$$F_{X_t}[X(t)] = F_X[X(t)] \quad (3)$$

2.6. MODELOS LINEALES AUTORREGRESIVOS O UNIVARIADOS

En un modelo autorregresivo lineal, el valor real de una variable se iguala a una suma ponderada de un número asignado previamente de valores ponderados y una variable que es completamente aleatoria, en este caso, la palabra lineal, significa que el valor real es dependiente aditivamente sobre valores pasados y no se refiere a sus cuadrados o raíces cuadradas.

Este modelo es la mejor aproximación a procesos naturales como los gastos en un río se complementan con cantidades de agua subterránea dependientes en el tiempo, escurrimiento superficial, retención en una cuenca y otros problemas parecidos.

El método para la solución de series de tiempo proviene del trabajo clásico de Markov (1907). La teoría básica de Yule (1921, 1927) fue quien evidentemente trabajó con estos modelos, fue quien uso por primera vez el término de correlación. Wold (1954) hizo un análisis teórico de estos modelos, posteriormente Box y Jenkins (1976) fueron quienes presentaron una metodología sistemática para el desarrollo de estos modelos (Kottogoda 1980 citado por Alvarado 2007).

El modelo aplicado, un componente estocástico Z_t , como el escurrimiento mensual, anual, etc., el cual se analiza como una variable aleatoria. Para esto se asume que la tendencia y la temporalidad son removidas con anterioridad. Se propone que el valor Z_t en el tiempo t se conforma de la suma ponderada de los p valores en el tiempo $t-1, t-2, \dots, t-p$ y un número aleatorio a_t ; los a_t valores $t=0, \pm 1, \pm 2, \dots$ son mutuamente independientes, siguen una distribución normal y están idénticamente distribuidos.

2.6.1. Modelo Autorregresivo AR (p)

Según Alvarado (2007) el modelo lineal autorregresivo de orden p o AR (p) tiene la forma:

$$z_t = \phi_{p,1}z_{t-1} + \phi_{p,2}z_{t-2} + \phi_{p,3}z_{t-3} + \dots + \phi_{p,p}z_{t-p} + a_t \quad (4)$$

$$= \sum_{i=1}^p \phi_{p,i}z_{t-i} + a_t \quad (5)$$

Donde $\theta_{p,i}$ $i=1, 2,3,\dots,p$ son los parámetros autorregresivos o ponderaciones. Las propiedades de Z_t y a_t son $E(Z_t) = E(a_t)=0$, $\text{var}(Z_t) = E(Z_t^2) = \sigma_a^2$, $\text{var}(a_t) = E(a_t^2) = \sigma_a^2$, $\rho_k = E(Z_t Z_{t-k}) / \sigma_a^2$ y para $k=1,2,3,\dots, E(a_t a_{t-1}) = E(a_t z_{t-k}) / \sigma_a^2$ y para $k = 1, 2,3,\dots, E(a_t a_{t-k}) = E(a_t Z_{t-k}) = 0$ donde la última igualdad significa que el número aleatorio actual es independiente o valor medio del término entre paréntesis y var es la varianza, ρ_k es el coeficiente de autocorrelación teórico y Z_t es la variable en estudio.

De ahora en adelante se asume, sin perder la generalidad, que la varianza σ_a^2 del componente estocástico a_t es igual a uno. Esto significa que, en las aplicaciones, los números generados con la ecuación anterior se deben multiplicar por la desviación estándar de la variable moderada y la media se debe sumar a cada número. Estos y otros parámetros se estiman a partir de secuencias de observaciones (Alvarado 2007).

2.6.2. Modelos de promedios móviles MA (q)

Alvarado (2007), menciona en el proceso de promedios móviles MA (q) de orden q, el valor de la variable aleatoria es una suma ponderada de q+1 números aleatorios. Aquí la ponderación asignada a la suma de los números aleatorios es la unidad, y los números aleatorios generados en tiempos anteriores se multiplican por $-\theta_{q,1}, -\theta_{q,2} \dots -\theta_{q,q}$ respectivamente. De acuerdo con esto el modelo tiene la forma:

$$z_t = a_t - \theta_{q,1} a_{t-1} - \theta_{q,2} a_{t-2} \dots - \theta_{q,q} a_{t-q} \quad (6)$$

$$z_t = a_t - \sum_{j=1}^q \theta_{q,j} a_{t-j} \quad (7)$$

Donde $\theta_{q,j}$ es el parámetro de promedios móviles. El componente estocástico a_t es equivalente en este caso al resultado de un filtro línea con ruido blanco debido a que el pronóstico de Z_t se hace a partir de errores aleatorios a_t . Tal como anteriormente $E(Z_t) = 0$ y $E(z_t^2) = 1$; también $E(a_t a_{t-k}) = 0$ para $k \neq 0$. Si se eleva al cuadrado la ecuación anterior y se toman esperanzas, la varianza y el componente aleatorio se estima con:

$$\sigma_a^2 = \frac{1}{1 + \theta_{q,1}^2 + \theta_{q,2}^2 + \dots + \theta_{q,q}^2} \quad (8)$$

Si se multiplica el lado izquierdo de la ecuación (8) por z_{t-k} y el lado derecho por $a_{t-k} - \theta_{q,1}a_{t-1-k} - \theta_{q,2}a_{t-2-k} - \dots - \theta_{q,q}a_{t-q-k}$ se toman esperanzas y se sustituyen por σ_a^2 en la ecuación $\Phi_p = P_p^{-1}p_p$ se tiene:

$$\rho_k = \frac{\theta_{q,k} + \theta_{q,1}\theta_{q,k+1} + \theta_{q,2}\theta_{q,k+2} + \theta_{q,1}\theta_{q,k+1} \dots \theta_{q,q}\theta_{q,q-k}}{(1 + \theta_{q,1}^2 + \theta_{q,2}^2 + \dots + \theta_{q,q}^2)} \quad (9)$$

y $\rho_k = 0$, $k > q$ donde ρ_k es el coeficiente teórico de autocorrelación del retraso k .

2.6.3. Estimación de los parámetros del modelo

Los métodos de estimación de parámetros generalmente se dividen en tres categorías: método de momentos, método de máxima verosimilitud y método de mínimos cuadrados.

El método de momentos se basa en tomar tantas ecuaciones de momento como el número de parámetros sustituyendo los momentos de población por los momentos de muestra y resolviendo las ecuaciones simultáneamente para los parámetros. En el método de máxima verosimilitud, primero se determina la función de verosimilitud (esto es el logaritmo) y los parámetros correspondientes a dicho máximo son los estimadores de máxima verosimilitud (Maidment 1993 citado por Alvarado 2007).

En el método de mínimos cuadrados, los parámetros que minimizan la suma de los residuos cuadrados $\sum \varepsilon_t^2$ son los estimadores de mínimos cuadrados; esto es particularmente útil para los modelos ARMA.

Los estimadores de momentos generalmente están disponibles para todos los modelos. El método de máxima verosimilitud es el método de estimación más eficiente en un sentido de error cuadrático medio, aunque los sesgos pueden ser un problema, especialmente para muestras pequeñas. El método de los momentos ha sido atractivo en la práctica, ya que es más fácil de aplicar y las correcciones por sesgo están disponibles para algunos modelos. El método de mínimos cuadrados es una aproximación al método de máxima verosimilitud, y generalmente ambos métodos requieren una solución numérica.

Si la ecuación (4) se multiplica por z_{t-1} y se toman esperanzas, se tiene que:

$$E(z_t z_{t-1}) = \Phi_{p,1}E(z_{t-1} z_{t-1}) + \Phi_{p,2}E(z_{t-2} z_{t-1}) + \Phi_{p,3}E(z_{t-3} z_{t-1}) + \dots + \Phi_{p,p}E(z_{t-p} z_{t-1}) + E(a_t z_{t-1}) \quad (10)$$

Dado que $E(a_t, z_{t-1}) = 0$ y tomando en cuenta otras propiedades que se dieron anteriormente, se obtiene la siguiente ecuación:

Donde $\phi_{p,p}$ es el parámetro auto regresivo, ρ_p es el coeficiente de autocorrelación teórico. Además, si la ecuación (4) se multiplica por $z_{t-2}, z_{t-3}, \dots, z_{t-p}$ y se toman esperanzas después de cada multiplicación, se obtienen un número de p relaciones llamadas Yule Walker (Alvarado 2017). Estas se pueden representar en forma matricial de la siguiente manera:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \cdot \\ \cdot \\ \cdot \\ \rho_p \end{bmatrix} = \begin{bmatrix} \rho_1 & \rho_1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_2 & \rho_2 & \rho_2 & \dots & \rho_{p-2} \\ \rho_3 & \rho_3 & \rho_3 & \dots & \rho_{p-3} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{p,1} \\ \phi_{p,2} \\ \phi_{p,3} \\ \cdot \\ \cdot \\ \cdot \\ \phi_{p,p} \end{bmatrix} \quad (11)$$

O abreviando se puede escribir:

$$\rho_p = P_p \phi_p \quad (12)$$

En donde una condición necesaria para que sea un proceso estacionario es que la matriz de auto correlación P_p sea positiva, esto es, el determinante y todos sus determinantes menores principales deben ser mayores o iguales a cero. Por lo tanto:

$$\phi_p = P_p^{-1} \rho_p \quad (13)$$

Para resolver la ecuación (13), los estimadores de los coeficientes de auto correlación $r_1, r_2, r_3 \dots r_p$ se sustituyen en P_p y ρ_p y con las ecuaciones de Yule-Walker para estimar los parámetros auto regresivos $\phi_{p,1}, \phi_{p,2}, \phi_{p,3} \dots \phi_{p,p}$. Para un proceso normal lineal, esto da una buena aproximación con los métodos de máxima verosimilitud o de mínimos cuadrados.

Además Maidment (1993) da a conocer que los estimadores $\hat{\mu}$, $\hat{\sigma}_e^2$ y $\hat{\phi}_1$ de momento del modelo AR (1) se pueden obtener de las ecuaciones:

$$\hat{\mu} = \hat{x} \quad (14)$$

$$\hat{\sigma}_s^2 = \frac{S^2}{(1 - r_1^2)} \quad (15)$$

$$\tilde{\theta}_1 = r_1 \quad (16)$$

En el que \hat{x} , s^2 y r_1 son la media muestral, la varianza y el coeficiente de autocorrelación lag-1. Un estimador de una estadística de población está sesgado cuando su valor medio es diferente de la estadística de población. El estimador r_k es un estimador sesgado (hacia abajo) de ρ_k . Se han sugerido varios procedimientos para corregir los valores de r_k de muestras limitadas. Wallis y O' Connell sugirieron la siguiente corrección para obtener un estimador imparcial de ρ_1 :

$$\hat{\rho}_1 = \frac{r_1 N + 1}{N - 4} \quad (17)$$

Donde N es el tamaño de la muestra. Igualmente, s^2 es un estimador sesgado de σ^2 si la serie es no correlacionada. Cuando la serie es auto correlacionada el tamaño efectivo de la muestra N es reducida. En este caso, un estimador de la varianza puede determinarse por:

$$\hat{\sigma}^2 = \frac{(N - 1)s^2}{N - K} \quad (18)$$

Donde:

$$K = \frac{[N(1 - \hat{\rho}_1)^2 - 2\hat{\rho}_1(1 - \hat{\rho}_1^N)]}{[N(1 - \hat{\rho}_1)^2]} \quad (19)$$

El estimador γ_0 es un estimador sesgado de γ si la serie s de tiempo es no correlacionado y distribuido normalmente. Para series no normales, propusieron el estimador insesgado:

$$\hat{\gamma}_o = \frac{Lg[A + Bg^2(L^2/N)]}{\sqrt{N}} \quad (20)$$

$$A = 1 + 6.5N^{-1} - 20.2N^{-2} \quad (21)$$

$$B = 1.48N^{-1} + 6.77N^{-2} \quad (22)$$

$$L = \frac{N - 2}{\sqrt{(N - 1)}} \quad (23)$$

Donde $\pm L$ es el límite teórico de asimetría de una muestra de tamaño N.

Mientras los estimadores de momento MA (1) se pueden obtener:

$$\hat{\theta}_1 = -\frac{1}{\theta_1} - \frac{1}{r_1} \quad (24)$$

$$\hat{\sigma}_e^2 = S^2(1 - r_1^2) \quad (25)$$

2.7. PRONÓSTICO

Sea t el periodo actual para el pronóstico, el principal interés es en periodos de tiempo futuro para a variable en series de tiempo, denotadas como Z_{t+l} donde $l \geq 1$. El periodo de tiempo t es llamado el origen del pronóstico, y l es llamado el pronóstico hacia delante o futuro. En el análisis de los modelos ARMA, el pronóstico depende de la observación de una variable realizada hasta el periodo t. Sea la información contenida en un conjunto de observaciones (Z_t, Z_{t-1}, \dots) , designado como I_t . Entonces el pronóstico de Z_{t+l} , será designado como $\hat{z}_t(l)$, esto es:

$$\hat{z}_t(l) = E(z_{t+l}/I_t) \quad (26)$$

Refiriéndose a un origen de tiempo t es deseable hacer un pronóstico con el error medio cuadrático mínimo de Z_{t+l} donde $l \geq 1$ es el intervalo de tiempo futuro. El pronóstico de un modelo ARMA (p, q) se puede expresar como una suma infinita ponderada de valores aleatorios previos al tiempo t (Alvarado 2007):

$$\hat{z}_t = \mu + \psi_0 a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \quad (27)$$

Donde ψ_i son los coeficientes de los valores aleatorios anteriores al tiempo t, $\hat{\mu}$ es la media de la serie histórica. El cálculo de ψ_i es:

$$\psi_0 = 1 \quad (28)$$

$$\psi_1 = \phi_1 - \theta_1 \quad (29)$$

$$\psi_2 = \phi_1(\phi_1 - \theta_1) \quad (30)$$

Y en general:

$$\psi_j = \phi_1^{j-1} (\phi_1 - \theta_1) \quad (31)$$

Así la varianza del error del pronóstico con ψ_0 es:

$$\text{var}[e_t(l)] = E[e_t^2(l)] - \sum_{j=0}^{l-1} \psi_j^2 \sigma_a \quad (32)$$

Donde $e_t(l)$ es el error en el origen t para el pronóstico en el periodo l, var es la varianza del error y σ_a es la varianza de las diferencias \hat{a}_t vistas como un proceso aleatorio.

El intervalo de confianza $(1 - \alpha)$ para un pronóstico con error medio cuadrático mínimo $z_t(l)$ es:

$$z_{t+l} = z_t(l) \pm u_{1/2} \left\{ 1 + \sum_{j=1}^{l-1} \psi_j^2 \right\}^{1/2} \sigma_a \quad (33)$$

Donde $u_{\alpha/2}$ es el nivel de probabilidad para una distribución normal estándar.

2.7.1. Simulación estocástica

Según Naylor (1975) citado por Aliaga (1980), la simulación es una técnica numérica para conducir experimentos en una computadora digital, las cuales requieren ciertos tipos de

modelos lógicos y matemáticos, que describen el comportamiento de un negocio o un sistema económico (o algún componente de ellos) en periodos extensos de tiempo real.

Los eventos de sequía, definido por el déficit y la duración observados de las series de tiempo hidrológicas generalmente son de un periodo corto, de manera que pueden ser bastante diferente aún en las series sintéticas, por lo que la preocupación es evaluar en qué medida un evento de magnitud extrema es representativo del periodo histórico.

La severidad de una sequía puede ser drásticamente cambiada sumando o restando un año al periodo de análisis. Para poder hacer cierta estimación de la probabilidad de ocurrencia de sequías severas, se simula el proceso estocástico para generar varias secuencias. Si la generación se realiza correctamente, se supone que las secuencias sintéticas pueden ocurrir probablemente en el futuro como pueden ser una repetición de la serie histórica.

Para simular el proceso estocástico hidrológico, una secuencia de números aleatorios independientes uniformemente distribuidos es primeramente generado; luego se genera números normales o log-normales según sea el mejor ajuste con sus respectivos parámetros, generándose números dependientes con el modelo Markovino seleccionado para finalmente, mediante un proceso de adición de las componentes determinísticas se obtenga las series hidrológicas sintéticas (Aliaga 1980).

2.7.2. Generando series sintéticas

La generación de datos es un tópico en hidrología estocástica y ha recibido bastante atención en la literatura hidrológica. La generación de datos es usada por hidrólogos para muchos propósitos. Estos incluyen, por ejemplo, tamaño del reservorio, planeamiento y manejo de un reservorio existente, y confiabilidad de un sistema hidráulico tal como un abastecimiento de agua o sistema de irrigación (Salas *et al.* 1985).

La generación de datos puede ayudar en la toma de decisiones clave, especialmente en situaciones críticas tales como periodos prolongados de sequía. La principal filosofía detrás de la generación de datos sintéticos es que las muestras sintéticas generadas preserven ciertas propiedades estadísticas que existen en los procesos hidrológicos naturales (Sveinsson *et al.* 2007).

Como un resultado, cada muestra generada y la muestra histórica no es más probable de ocurrir que cualquiera de las muestras históricas.

La generación de series de tiempo sintéticas está basada en los modelos de aproximación y esquemas. Una vez que el modelo ha sido definido y los parámetros han sido estimados, uno puede generar muestras sintéticas basadas en el modelo. SAMS permite generar datos sintéticos y eventualmente comparar características importantes de los datos históricos y los datos generados. Tal comparación es importante para chequear si el modelo usado en la generación es adecuado o no. Si las estadísticas históricas y generadas importantes son comparables, luego uno puede sostener que el modelo es adecuado.

Las series hidrológicas históricas generalmente son de un periodo corto y los eventos sean máximo o mínimo, pueden ser bastante diferente aún en las series largas, por lo que la preocupación es evaluar en qué medida un evento de magnitud extrema es representativo del periodo histórico.

2.8. SEQUÍA

De acuerdo con Vega (2018) no existe una definición universalmente aceptable de la sequía debido a la compleja naturaleza del fenómeno, encontrándose más de 150 definiciones para las sequías.

Según la disciplina científica desde la que se analiza al fenómeno, la definición de sequía depende de cada caso específico en que se utilice, variando con el espacio y, el tiempo, significando aspectos para cada uno de sus usos y para cada persona. (Aliaga 1980). Encontrar una definición de sequía requiere conocer sus diferentes formas y características como es percibida dentro de un cierto tipo de actividad, tales como agrícolas, económicas, sociales, y sobre todo natural desde un punto de vista meteorológico e hidrológico.

Un periodo prolongado, una estación, un año o varios años de lluvia deficiente en relación con el promedio estadístico para una región definen una sequía (Schneider *et al.* 2011), cuales condiciones meteorológicas anormales secas, para que la falta de precipitación cause un grave desequilibrio hidrológico (OMM 1992) o un agudo desequilibrio hídrico que perjudica los sistemas de producción del recurso tierra (CLD 2004).

La sequía es un fenómeno natural que se extiende de manera irregular a través del tiempo y el espacio, tiene la particularidad de ser de lento desarrollo temporal y amplia cobertura espacial, su ocurrencia no provoca el colapso ni la falla destructiva de la infraestructura; sin embargo, puede condicionar las operaciones de los sistemas de recursos hídricos

imponiendo muchas veces condiciones más exigentes que las de diseño con graves consecuencias económicas y sociales (Fernández 1997).

Tal disminución o déficit temporal en la disponibilidad de agua natural, representa un peligro natural (Vogot y Somma 2000), encontradas por actividades productivas naturales o del hombre como el abastecimiento de agua, la agricultura, climatológica, social económica o particularmente hidrológica desde un enfoque operacional de embalses orientado al estudio de estas Linsley *et al.* (1949), impactan su desarrollo o función, y a menudo exacerbado por el hombre y su sociedad.

2.8.1. Tipos de sequías

Wilhite y Glantz (1985) citados por Vega (2018), agruparon diversas definiciones de sequías, en cuatro grupos según la disciplina científica desde la cual se analiza el fenómeno: sequía meteorológica, agrícola, hidrológica y socioeconómica.

a. Sequía meteorológica o climatológica

Relacionada con la disminución de las precipitaciones, siendo evaluada mediante la comparación ente la precipitación del periodo de análisis y un indicador estadístico de la serie temporal que generalmente es la media. Existen diversas definiciones de la sequía meteorológica que han sido desarrolladas para regiones específicas, debido a que las variaciones en los regímenes de precipitación tienen una fuerte influencia de las características climáticas de cada región.

b. Sequía edáfica o agrícola

Se expresa en función de la deficiencia de humedad en el suelo disponible para las plantas, en muchos casos es valorada de acuerdo con los daños provocados sobre la producción agrícola. Debido a que la cantidad de agua es diferente para cada cultivo, y varía a lo largo de las fases de crecimiento de los cultivos, es complejo definir umbrales de sequía incluso para una misma región. Este tipo de sequía no solo depende de las condiciones meteorológica sino también de las características biológicas del cultivo y las propiedades del suelo. Por lo que no es equivalente a la sequía meteorológica. Si los niveles de humedad del suelo son suficientes para sustentar el crecimiento de un determinado cultivo durante el periodo que se presente sequía meteorológica, la sequía agrícola no se producirá.

c. Sequía hidrológica

Se asocia a la deficiencia en el caudal o volumen de aguas superficiales y subterráneas, así como el volumen de almacenamiento. Las sequías hidrológicas tienen un desfase entre la ocurrencia de las sequías meteorológicas o agrícola, puesto que se requiere de un desfase temporal para que se propaguen las anomalías pluviométricas a los diversos componentes del ciclo hidrológico.

A diferencia de la sequía agrícola que tiene lugar al poco tiempo de presentarse la sequía meteorológica, la sequía hidrológica puede tardar de meses a años dependiendo de la región desde el inicio de la sequía meteorológica; y si las lluvias retornan en poco tiempo, la sequía hidrológica no llegará a manifestarse.

d. Sequía socio económica

Se produce cuando la disponibilidad de agua disminuye hasta el punto de generar daños económicos o sociales sobre poblaciones de zonas afectadas por la escasez de precipitaciones.

2.8.2. Antecedentes de sequías en la zona de estudio

Las sequías hidrológicas observadas del río Chancay Lambayeque han sido reportadas principalmente por el Operador del Sistema Hidráulico Tinajones del Proyecto Especial Olmos Tinajones, la Entidad Prestadora de Servicio de agua potable, Junta de Usuarios de la Infraestructura Hidráulica, la Autoridad Nacional del Agua y el Servicio Nacional de Meteorología e Hidrología, estos a través de diversos estudios, planes de gestión, monitoreos o avisos hidrológicos y medios de comunicación escrita.

La región costera por su situación geográfica próxima al Ecuador con influencia en su temperatura debido a la circulación atmosférica del anticiclón del pacífico, la presencia de la cordillera de los andes que frena el intercambio de masas de aire entre la zona tropical húmeda del oeste y la zona árida costera, la formación y persistencia de la inversión térmica por el anticiclón subtropical de la zona del pacífico sur como la corriente oceánica de Humboldt determinan condiciones meso y micro meteorológicas muy variadas, extremadamente secas con precipitaciones entre los 50 y 200 mm con alteraciones del

clima en el régimen pluviométrico influenciadas por la variabilidad del El Niño con lluvias y caudales por encima de lo normal o en un contexto no Niño con reducidas lluvias como las de 1990 y 1983 con déficit de caudales de los ríos (SENAMHI, s.f.).

Los ríos de la vertiente del pacífico como del río Chancay Lambayeque son de corto recorrido entre 90 y 200 Km hasta el pacífico, de tal manera el escurrimiento sucede con prontitud a comparación con los 130 días para que se originen las precipitaciones.

Por tal motivo, la masa descargada en época de lluvias es el 74% del volumen total recibidos en las cuencas, hecho que resulta una disminución en las descargas de los meses siguientes. De este modo, las aguas presentan dos periodos distintos de cuatro meses, uno de aguas medias (17% de la masa anual) y uno con sequías (9% de la masa total) a falta o ausencia de lluvias en la cuenca (SENAMHI, s.f.).

La variabilidad interanual asociada a El Niño explica en buen porcentaje las deficiencias de lluvias en los andes, principalmente en la sierra central y sur occidental. Esta relación no es absoluta, puesto que se ha evidenciado casos de sequías intensas durante años ENSO de fase neutral, tal como el ocurrido durante el verano del año 1990. La predominancia del patrón típico del ENSO positivo (niño) para el verano (enero, febrero, marzo) de los años más secos (1982,83,87,88,90,92,04,05,16) en el cual revela una clara asociación con el menor potencial para la convección en los Andes y Norte de Sudamérica, no así para la región suroriental (SENAMHI 2019b).

En la zona de estudio la sequía conocida por falta de lluvias tiene la acepción de falta o depresión de los niveles de agua superficial del río Chancay-Lambayeque debajo de los mínimos de consumo agrícola, poblacional, etc. Por ejemplo, varios autores SENAMHI (s.f.), SENAMHI (2019a), PEOT (2004), ANA (2013) y Vega (2020) informan de los sucesos de sequías extrema y moderada acontecidos los años 1963, 1968, 1982,1985,1986,1990, 1991, 1997, 2004 y 2014 donde las descargas del río Chancay Lambayeque disminuyeron considerablemente durante el primer trimestre.

El SENAMHI, a través del índice estandarizado de precipitación SPI identifico cinco episodios moderadamente secos y extremadamente secos de tres meses consecutivos en el departamento de Lambayeque (Cuadro 1).

Cuadro 1. Características de las sequías en términos de intensidad y severidad.

Año	Intensidad	Severidad	Categoría
1982	-2.3	5.2	Extremadamente seco
1985	-2.0	9.5	Extremadamente seco
1986	-1.2	1.2	Moderadamente seco
1990	-2.1	29.3	Extremadamente seco
1991	-1.3	1.3	Moderadamente seco

Adaptado de Caracterización Espacio Temporal de la Sequía en los Departamentos Alto Andinos del Perú (1981-2018), SENAMHI (2019b).

Por ejemplo, las aguas que caen en forma de precipitación en las cabeceras de los valles de la región costera han sido durante 3 o 4 meses del año SENAMHI (s.f.) en tanto el río Lambayeque presentó anomalías de -37% respecto al caudal histórico desde noviembre 2018 a enero 2019, a consecuencia del déficit de precipitación respecto a la normal de menos 15% a 15% (SENAMHI 2019b).

Sosa (2016) del compendio estadístico agrario 50/91, da cuenta durante la campaña agrícola 1967-1990 los eventos de sequías hidrológicas fueron de intensidad débil (Cuadro 2).

Cuadro 2. Eventos climáticos en el Perú 1950-1998.

Campaña	Evento	Intensidad
67/68	Sequía	Débil
68/69	Sequía	Débil
69/70	Sequía	Débil
77/78	Sequía	Débil
78/79	Sequía	Débil
89/90	Sequía	Débil

Fuente: Adaptado de análisis de la sequía hidrológica en el Perú, Sosa (2016).

Sosa (2016) ha comprobado la relación entre la sequía hidrológica y la sequía de lluvia del análisis del caudal y lluvia (1973-2013) en la cuenca del río Chancay Lambayeque, destacando condiciones de sequía de lluvia e hidrológica el periodo 1997-1998 como condiciones sin sequía de lluvia e hidrológica los periodos 1982-1983 y 1989-1990.

También del análisis de riesgo de sequía hidrológica en los episodios ENOS en temporada del año hidrológico, periodo 1973-2013, el riesgo de sequía en la estación Racarumi fue 1.5% (Niño) y 0.20% (Niña).

Los impactos a consecuencia de las sequías hidrológicas en la zona de estudio no se precisan según el origen del agua, el uso y lugar de aprovechamiento, la ANA (2013) a través del plan participativo de gestión de recursos hídricos de la cuenca Chancay Lambayeque, da cuenta el impacto de las sequías afectaron a 109 660 has de riego, mientras el SENAMHI (s.f.) menciona la sequía del 67-68 motivó una reducción de cultivo de 125 068 hectáreas.

2.8.3. Propiedades de la sequía

Una sequía puede ser caracterizada por su duración, déficit de agua acumulada (o déficit promedio o intensidad), terminación y extensión espacial. (Aliaga 1985 y Goldman 1985).

Estas características pueden ser determinadas considerando una serie abastecimiento de agua denotada por $x_t, t = 1, 2, \dots$ y un umbral de demanda de agua constante x_0 .

De acuerdo con Salas *et al.* (2005), con la definición de sequía sugerida por Yevjevich (1967), un evento de sequía es tomado como una sucesión de períodos consecutivos (run) en que el suministro de agua x_t permanece bajo x_0 , donde el run es precedida y sucedida por el suministro de agua que es igual o más grande que x_0 . De esta forma, la extensión de sequía L (o extensión run negativa) es el número de intervalos de tiempo consecutivos (por ejemplo, años) en que $x_t < x_0$, precedida y seguida por (al menos un período donde) $x_t \geq x_0$. Esta definición de sequía es esquemáticamente mostrada en la (Figura 1).

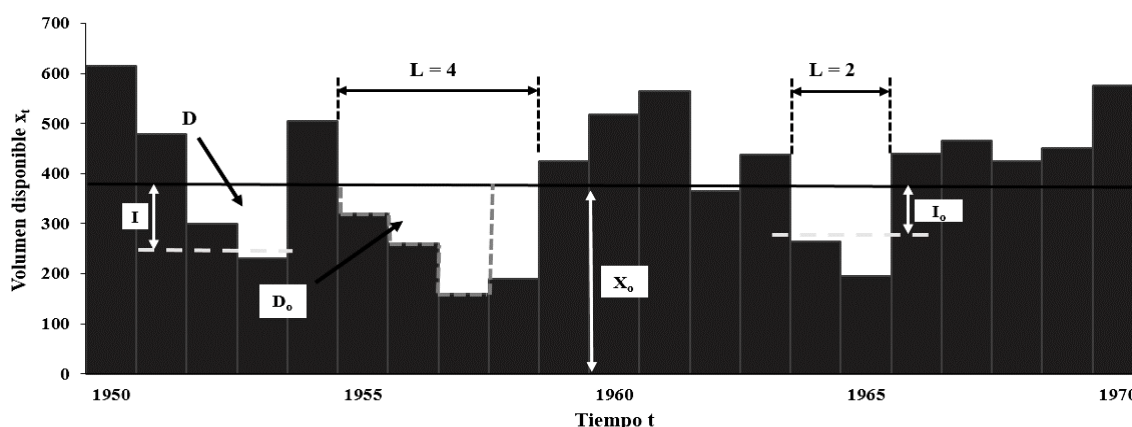


Figura 1. Suministro de agua x_t , demanda de agua x_0 , y propiedades de la sequía: (a) duración L ; (b) déficit acumulado D ; (c) intensidad I .

Fuente: Salas, *et al.* 2005.

Además, el déficit acumulado es el déficit por toda la duración de la sequía, por ejemplo:

$$D = \sum_{j=t}^{t+L-1} (x_0 - x_j) = \sum_{j=t}^{t+L-1} d_j \quad (34)$$

y la intensidad de sequía es el déficit medio sobre la duración de la sequía, por ejemplo $I = D/L$. Además, D_0 representa alguna cantidad dada de déficit donde $D_0 \geq 0$, y I_0 representa algún valor de la intensidad donde $I_0 > 0$.

2.8.4. Identificación de la sequía

Un problema mayor en el análisis de sequías es la separación de sus ocurrencias desde el registro hidrológico. La dificultad está en el hecho que la ocurrencia de las sequías depende de la interacción entre la ocurrencia natural del agua (factores hidrometeorológicos) y la intención de uso del agua (uso operacional). Como un ejemplo de esta dificultad, considera la percepción de la sequía desde el punto de vista del hidrólogo como un caudal debajo de la normal.

La percepción de sequía varia, considerando la variabilidad regional de sequías meteorológicas, Dracup *et al.* (1980) reportaron periodos de sequías están considerados ocurran después de seis días sin lluvias en Bali y después de dos años sin lluvias en Libia. La intención del uso del agua es un factor crítico en sequías hidrológicas. Como Beard y Kubik (1972) puntualizaron, los caudales que están considerablemente debajo de lo normal para periodos cortos (sequías intensas de corta duración) puede ser muy significativo en áreas donde la demanda es una pequeña fracción de la normal disponibilidad, pero de poca significancia donde un amplio almacenamiento está presente. Por otra parte, largos periodos de caudales ligeramente debajo la media (duraciones largas de intensidad baja) puede ser significativo para usos que dependen del almacenamiento, pero de poca importancia a una fracción pequeña de usuarios.

Así, la definición de sequías depende fuertemente del particular enfoque del análisis. Una sencilla caracterización de este fenómeno no es posible. En vista de las diferentes percepciones de sequía, Dracup *et al.* (1980) propusieron cuatro tareas para el comienzo de su análisis.

La primera tarea es determinar la naturaleza del déficit hídrico y se refiere a la elección de analizar cualquiera de los caudales con la distinción entre la causa de la sequía (caudales) y el impacto debido a la sequía (suministro de agua). De tal manera, para los hidrólogos/planificadores el principal interés está en el impacto. Por esta razón, el análisis está restringido a un déficit individual tal como un caudal, sin embargo, ya sea el enfoque combinado o individual es válido (Goldman 1985).

La segunda tarea es identificar el periodo de tiempo integro para el análisis. El periodo de tiempo integro es el incremento del tiempo; horas, días, meses, estaciones, años, etc., sobre el cual la data hidrológica es promediada en el análisis de sequía, y es uno de los dos factores que determinan el número de eventos de sequía en el registro de hidrológico (el nivel de truncamiento es otro factor). Un efecto obvio del incremento de la longitud del periodo integro es la pérdida correspondiente de la información acerca del proceso hidrológico. Por ejemplo, caudales estacionales los cuales son sucesivamente más altos o bajos que la normal no necesariamente son reconocidos cuando empleamos un periodo integral en el análisis (Goldman 1985).

La elección del periodo integral distingue entre las definiciones generalmente aceptando la definición de valores de caudales extremos, caudales máximos o caudales mínimos sobre los más bajo del espectro. Las inundaciones y caudales bajos son generalmente considerados medidos instantáneamente de los caudales. Por ejemplo, una inundación es descrita en términos de una descarga pico, dice las inundaciones de 100 años. La descarga pico de una inundación de 100 años tiene un porcentaje de oportunidad de ser igualado o excedido en algún año dado.

Una tercera tarea es establecer el nivel de truncamiento empleado para distinguir los eventos de sequía del registro hidrológico. El nivel de truncamiento refleja la demanda socio económica del suministro de agua disponible. Por ejemplo, el caudal medio anual (o alguna fracción de la media) puede ser usado para representar la demanda esperada de una municipalidad del caudal disponible. Sin embargo, la demanda no necesariamente es constante y puede ser representada algunas veces variando los niveles de truncamiento. La suposición es hecha para la discusión (a no ser lo contrario) que el nivel de truncamiento sea conocido, y para simplificar el motivo es resuelto como un valor constante (Goldman 1985).

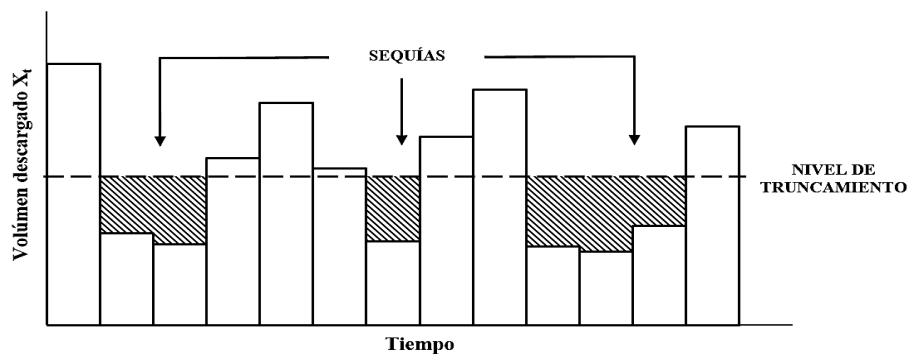


Figura 2. Nivel de truncamiento.

Fuente: Goldman (1985).

Como puede ser visto de la Figura 2, periodos de bajo caudal del nivel de truncamiento son identificados como una sequía o periodos caudales bajos y caudales por encima del nivel de truncamiento como un periodo de caudales altos o crecientes. En efecto, la separación (y simetría) entre la definición de caudal bajo, sequía, caudal alto e inundación puede verse con bastante facilidad en la Figura 3 por combinación los conceptos de periodo integral y nivel de truncamiento.



Figura 3. Caudal máximo, caudal mínimo, sequía, inundación y el periodo integral.

Fuente: Goldman (1985).

La delineación del área de estudio se basa de ambos factores de homogeneidad, estadísticas y geomorfológico. Los factores geomorfológicos que delinean un área incluyen la topografía (condiciones de montaña son un factor obvio), almacenamiento local (lagos) y propiedades del suelo.

En el enfoque estadístico, los lugares son agrupados basados en estadísticas similares del registro hidrológico. Esta vez la delineación de una extensa área de una sequía basada en estos métodos no han sido adecuadamente indagadas y es un área de investigación necesaria.

Goldman (1985) concluye, una sequía puede identificarse a partir de un registro hidrológico basándose en las tareas de análisis de sequía. Por lo tanto, el principal problema para abordarlo es construir un modelo de los procesos hidrológicos como ejemplo un modelo de caudales para predecir, o al menos estimar, la severidad potencial de sequías futuras, asumiendo que el periodo integral y nivel de truncamiento son especificados.

2.8.5. Análisis de los eventos de sequía

De acuerdo con la investigación realizada por Salas *et al.* (2005), en el análisis de la severidad de sequías y los riesgos asociados, se necesita especificar el evento de sequía bajo consideración. Por ejemplo, uno puede considerar solamente la duración de una sequía indiferente del déficit, la duración con un cierto grado de déficit, o la duración y una intensidad dada. Más específicamente, los siguientes casos son considerados.

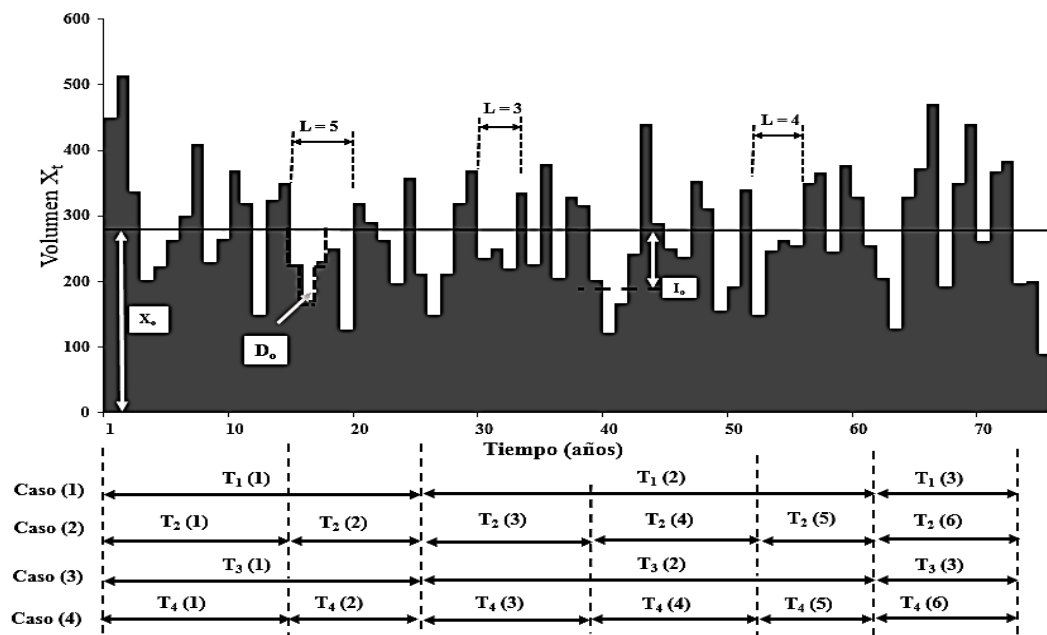


Figura 4. Definición esquemática de los tiempos de retorno $T_i(j), j = 1, 2, \dots$ de cuatro tipos de eventos de sequía: (1) $\{D > D_0 \text{ y } L = 3\}$; (2) $\{D > D_0 \text{ y } L \geq 3\}$; (4) $\{I \geq I_0 \text{ y } L \geq 3\}$; (3) $\{I > I_0 \text{ y } L = 3\}$.

Fuente: Salas, *et al.* 2005.

1.-Eventos de sequía con déficit acumulado D más grande que una cantidad especificada D_0 y duración L igual a un número dado de años l_0 , por ejemplo $\{D > D_0 \text{ y } L = l_0 \text{ (} l_0 =$

1,2,3...}). Las sequías donde $D_0 = 0$ son casos particulares donde solamente la duración especificada l_0 es considerada indiferente de la cantidad de déficit. La figura 4 muestra tres sequías con duración $L = 3$ años donde $D > D_0$ y los tiempos de recurrencia son T_1 (1), T_1 (2), y T_1 (3).

2.-Eventos de sequía donde $D > D_0$ y duración $L \geq l_0$, por ejemplo, $\{D > D_0 \text{ y } L \geq l_0 (l_0 = 1,2, \dots)\}$ Las sequías donde $D > D_0$ y $L \geq 1$ representan sequías indiferentes de la duración de la sequía; recíprocamente, las sequías donde $D > 0$ y $L \geq l_0$ representa sequías indiferentes de la cantidad de déficit. La figura 4 muestra seis sequías donde $D > D_0$ y $L \geq 3$ años, donde los tiempos de recurrencia son T_2 (1), T_2 (2), T_2 (3), T_2 (4), T_2 (5), y T_2 (6). Este tipo de sequía ocurre más frecuentemente que aquellos considerados en el caso 1 porque ellas incluyen duraciones de sequía más allá del umbral l_0 .

3.-Eventos de sequía con intensidad I más grande que una cantidad especificada I_0 y duración $L = l_0$, por ejemplo $\{I > I_0 \text{ y } L = l_0, (l_0 = 1,2, \dots)\}$. Como en el caso 1, las sequías donde $I_0 = 0$ es un caso particular donde solamente la duración de la sequía es considerada y es diferente de la intensidad. La fig. 4 muestra cuatro sequías con duraciones $L = 3$ años donde $I > I_0$ y los tiempos de recurrencia son T_3 (1), T_3 (2), T_3 (3), y T_3 (4).

4.-Los eventos de sequía donde $I > I_0$ y duración $L \geq l_0$, por ejemplo, $\{I > I_0 \text{ y } L \geq l_0, (l_0 = 1,2, \dots)\}$. Eventos donde $I > I_0$ y $L \geq 1$, representan sequías indiferentes de la duración de la sequía (no consideran la duración de la sequía), mientras $I > 0$ y $L \geq l_0$ representan sequías indiferentes (sin considerar) de la intensidad.

La Figura 4 muestra cinco sequías con $L \geq 3$ años y $I > I_0$ y los tiempos de recurrencia son T_4 (1), T_4 (2), T_4 (3), T_4 (4) y, T_4 (5).

La severidad de los cuatro tipos de eventos de sequía aquí definidos puede ser de interés para sistemas dependientes de almacenamiento de agua. Por ejemplo, sistemas de almacenamiento pueden ser diseñados para sustentar una sequía de duración especificada l_0 teniendo el déficit acumulado más grande, es decir D_0 , en un horizonte de planeamiento dado (por ejemplo 50 años). De esta forma, la severidad de sequías dada por $(D > D_0 \text{ y } L = l_0)$ o $(D > D_0 \text{ y } L \geq l_0)$ puede ser de interés. También, la severidad de este tipo de sequía

puede ser útil no solamente para administradores del agua, sino para el público en general como también en casos donde una sequía en curso puede haber ya transcurrido por l_0 años y puede tener ya un déficit igual a D_0 . Argumentos y ejemplos similares como los anteriores pueden ser dados para eventos de sequía en términos de intensidad, por ejemplo ($I > I_0$ y $L = l_0$) o ($I > I_0$ y $L \geq l_0$). Además, los casos donde ($L=1$ y $I > I_0$) son sequías anuales simples que pueden ser de interés para sistemas de abastecimientos de agua que cuenten principalmente con desvíos directos del río.

Dependiendo de la magnitud del déficit de agua D_0 (o intensidad I_0), la sequía puede no afectar adversamente aquellos usuarios de agua con derechos de agua así como compararlos con aquellos que tienen más derechos de agua en el sistema. Durante sequías extremas, aquellos con derechos de agua junior a menudo son incapaces de desviar agua. En cualquier caso, la severidad de varios tipos de eventos de sequía, así como los definidos aquí puede ser cuantificados por la estimación de sus períodos de retorno correspondientes.

2.9. FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD DE EVENTOS DE SEQUÍA

Según Guven (1983) y Shiau y Shen (2001) la función de distribución de probabilidad bivariada (pdf) de déficit acumulado D y extensión L puede ser expresado como:

$$f_{D,L}(d,l) = f_{\frac{D}{L}}(d)f_L(l) \quad (35)$$

Donde $f_{\frac{D}{L}}(d)$ es la función de distribución de probabilidad condicional de D dado $L = l$ y $f_L(l)$ es la función de distribución marginal de L . Deducir la expresión exacta de $f_{D,L}(d)$ considerando todas las características de los procesos fundamentales no es una tarea fácil. El déficit acumulado D es una suma de un número aleatorio de variable aleatorias; por ejemplo, D es la suma del déficit sucesivos $d_j = x_0 - x_j$ donde L es una variable aleatoria. Claramente los déficits simples d_t son auto correlacionados, pero como una aproximación asumimos que son variables gamma distribuidas idénticamente independientes (iid) con parámetro forma r y parámetro escala b , por ejemplo, $G(r,\beta)$. Luego sigue que D/L , que es la suma de $L = l$ es también una variante de gamma, $G(lr,\beta)$ según Guven (1983); Shiau y Shen (2001).

$$f_{\frac{D}{L}}(d) = \frac{1}{\beta \Gamma(lr)} \left(\frac{d}{\beta}\right)^{lr-1} \exp\left(-\frac{d}{\beta}\right) \quad (36)$$

Además, si denotamos la ocurrencia de un superávit (sobrante) por 1 y la de un déficit por 0 y asumimos que la secuencia de superávit y déficit sigue una Cadena de Márkov Simple, puede ser demostrado que la función de distribución de probabilidad de extensión de sequía L es geométrica. Por ejemplo:

$$f_l(l) = P_{01}(1 - P_{01})^{l-1} \quad (37)$$

Donde p_{01} representa la probabilidad transición de un superávit en el tiempo t dado que un déficit ocurrido en el tiempo t-1. De estos resultados, la función de distribución de probabilidad bivariado del déficit y la extensión se vuelve:

$$f_{D,L}(d, l) = \frac{l}{\beta \Gamma(lr)} \left(\frac{d}{\beta}\right)^{lr-1} \exp\left(-\frac{d}{\beta}\right) P_{01}(1 - P_{01})^{l-1} \quad (38)$$

Claramente debido a que D es continua y L es una variable aleatoria discreta, la distribución combinada dada por la ecuación (38) es mixta. Las distribuciones marginales de D y L pueden ser obtenidas de la ecuación (38) por integración. Por ejemplo, la función de distribución de probabilidad de D puede ser obtenida por:

$$f_D(d) = \sum_{l=1}^{\infty} f_{D,L}(d, l) = \sum_{l=1}^{\infty} \frac{l}{\beta \Gamma(lr)} \left(\frac{d}{\beta}\right)^{lr-1} \exp\left(-\frac{d}{\beta}\right) P_{01}(1 - P_{01})^{l-1} \quad (39)$$

Aunque la expresión anterior de la función de distribución de probabilidad no es escrita en forma cerrada, puede ser demostrado que se complica con dos requerimientos de una función de distribución de probabilidad; por ejemplo, es más grande o igual a cero sobre el rango de valores de d y se integra a 1 para los valores de d recorriendo desde 0 a ∞ . Además, desde la distribución condicional $f_{D/L}(d)$, ecuación 37, la distribución condicional de la intensidad de sequía I dada la extensión L puede ser derivada. Debido a que $I = D/L$, la condicional de la función de distribución de probabilidad de intensidad I dada $L = l$ se hace:

$$f_{\frac{l}{L}}(i) = \frac{l}{\beta\Gamma(lr)} \left(\frac{li}{\beta}\right) \exp\left(-\frac{li}{\beta}\right) P_{01}(1 - P_{01})^{l-1} \quad (40)$$

que es también gamma distribuido, por ejemplo, $G(lr, b/l)$

De esta forma, la función de distribución de probabilidad bivariado de intensidad y longitud puede ser encontrada en una forma similar como la ecuación 1. Como sigue:

$$f_{\frac{l}{L}}(i, l) = \frac{l}{\beta\Gamma(lr)} \left(-\frac{li}{\beta}\right) P_{01}(1 - P_{01})^{l-1} \quad (41)$$

Por lo tanto, por integración la función de distribución de probabilidad bivariado, la probabilidad de ocurrencia de varios eventos de sequía puede ser encontradas. En particular con referencia a los cuatro eventos de sequía (E) considerados previamente, las siguientes expresiones pueden ser encontradas:

P [D>D₀ y L = l₀]:

$$\int_{D_0}^{\infty} f_{D,L}(Z, l_0) = \int_{D_0}^{\infty} \frac{1}{\beta\Gamma(lr)} \left(\frac{z}{\beta}\right)^{l_0 r - 1} e^{-z/\beta} P_{01}(1 - P_{01})^{l_0 - 1} dz \quad (42)$$

P [D>D₀, L >= l₀]:

$$\int_{D_0}^{\infty} \sum_{l=1}^{\infty} f_{D,L}(z, l) dz ; \int_{D_0}^{\infty} \sum_{l=l_0}^{\infty} \frac{1}{\beta\Gamma(lr)} \left(\frac{z}{\beta}\right)^{lr - 1} e^{-z/\beta} P_{01}(1 - P_{01})^{l-1} dz \quad (43)$$

P [I> I₀, L = l₀]:

$$\int_{I_0}^{\infty} \sum_{l=l_0}^{\infty} f_{I,L}(z, l) dz ; \int_{I_0}^{\infty} \sum_{l=l_0}^{\infty} \frac{l}{\beta\Gamma(lr)} \left(\frac{lz}{\beta}\right)^{lr - 1} e^{-lz/\beta} P_{01}(1 - P_{01})^{l-1} dz \quad (44)$$

P [I> I₀, L>= l₀]:

$$\int_{I_0}^{\infty} \sum_{l=l_0}^{\infty} \frac{l}{\beta\Gamma(lr)} \left(\frac{lz}{\beta}\right)^{lr - 1} e^{-lz/\beta} P_{01}(1 - P_{01})^{l-1} dz \quad (45)$$

La distribución marginal de eventos de sequía $D > D_0$ o $L > L_0$ puede ser obtenida de ecuaciones 43 o 45, respectivamente, por dejar $l_0 = 1$. También notar que, a pesar de la aparente complejidad de las ecuaciones, las integraciones pueden ser realizadas eficientemente usando herramientas numéricas para los pdf gamma que están disponibles en la mayoría del software matemático y estadísticos.

2.9.1. Severidad y riesgo de sequía

Shiau y Shen (2001) proponen que, la severidad de sequía sea medida en términos de períodos de retorno o tiempo de recurrencia del evento de sequía particular bajo consideración. Se define la severidad de sequías siguiendo conceptos similares a aquellos comúnmente usados con relación a la severidad de inundaciones. Por ejemplo, cuando evaluamos la severidad de una sequía Q_0 que ha ocurrido en un río dado, uno comúnmente determina el período de retorno T del evento inundación $Q > Q_0$. Si la inundación tiene un período de retorno de 10 años; significa que uno esperará inundaciones más grandes que Q_0 es 10 años. Tal inundación no es una inundación rara.

Para determinar el período de retorno (o tiempo de recurrencias medio) de sequías, uno podría usar algunos eventos y escenarios de sequía E que son funciones de ambos déficit y duración de sequía (o intensidad) como lo describió Fernández y Salas (1999a) por ejemplo, $E = \{D > D_0 \text{ y } L = l_0\}$.

Una estimación del período de retorno del evento de sequía E es dado por el promedio del tiempo de recurrencia como:

$$T = \frac{1}{N_E} \sum_{j=1}^{N_E} T_E(j) \quad (46)$$

En que $T_E(j)$ denota el tiempo de recurrencia de dos eventos de sequía sucesivos E ; y $N_E =$ número de tales eventos de sequía. La Eq. 46 es particularmente útil para determinar los períodos de retorno de eventos de sequía basados en una muestra histórica o una muestra generada de un modelo estocástico.

Para deducir una expresión analítica para T , la fórmula dada para eventos de sequía definidos basados en el déficit acumulado será extendida para los casos de eventos de sequía definidos (conjuntamente) en términos de cualquier déficit acumulado o intensidad y extensión. El tiempo interarrival T_E entre dos eventos de sequía E (por ejemplo, $E = \{D > D_0 \text{ y } L = l_0\}$). Es:

$$T_E = \sum_{j=1}^{N_d} T(j) \quad (47)$$

Donde $T(j)$ = tiempo interarrival entre algunas dos sequías en general (por ejemplo, sequías no necesariamente caracterizadas por E ; y N_d = número de sequías hasta que el siguiente evento de sequía E ocurra. Luego, de la ecuación (47) el período de retorno T es el valor esperado de T_E o:

$$T = E[T_E] = E(N_d)E[(T(j))] \quad (48)$$

El valor esperado de N_d puede ser obtenido por observación porque los eventos de sequía específicos son independientes, lo que se mantiene igual en el caso que la serie fundamental tiene dependencia Markoviana, el número de sequías N_d tiene una distribución geométrica con parámetro dado por la probabilidad de el evento de sequía E . Luego su valor esperado es:

$$E(N_d) = \frac{1}{P(E)} \quad (49)$$

Donde $P(E)$ puede ser determinado de las ecuaciones (42) y (45), dependiendo del caso. De otra forma, el valor esperado del tiempo interarrival entre sequías, $T(j)$, es la suma de los de los valores esperados de extensión de sequía y extensión de superávit (no sequía). El primero se infiere (deduce) directamente de la ecuación (38) como $1/p_{01}$, mientras que el último puede ser expresado como $1/p_{10}$ simplemente por cambio de las definiciones de déficit y superávit. Se deduce que:

$$E[T(j)] = \frac{1}{P_{01}} + \frac{1}{P_{10}} = \frac{P_{01} + P_{10}}{P_{01} P_{10}} \quad (50)$$

Combinando las ecuaciones (48) y (50), el período de retorno de un evento de sequía E es obtenido como:

$$T = \frac{P_{01} + P_{10}}{P_{01} P_{10}} \frac{1}{P(E)} \quad (51)$$

Por ejemplo, el período de retorno de un evento de sequía $E = \{D > D_0 \text{ y } L = l_0 (l_0 = 1, 2)\}$ puede ser expresado como:

$$T = \frac{P_{01} + P_{10}}{P_{01} P_{10}} * \frac{1}{P[D > D_0, L = l_0]} \quad (52)$$

Además, en el caso que el proceso es independiente (por ejemplo, una secuencia de ensayos Bernoulli independientes), puede ser demostrado que la Eq. (51) se hace:

$$T = \frac{1}{p_1 p_0} * \frac{1}{P(E)} \quad (53)$$

Donde p_1 y p_0 son las probabilidades marginales.

El riesgo que un evento de sequía particular pueda ocurrir en un período de tiempo dado n , puede también ser de interés en el planeamiento de sistemas de recursos de agua. Por ejemplo, si un sistema de abastecimiento ha sido diseñado para cubrir con sequías definidas por $E = \{D > D_0 \text{ y } L = l_0\}$ puedan ocurrir durante el período de tiempo n . Tal riesgo puede ser determinado de muestras históricas o de muestras generadas como las ilustradas en lo futuro.

2.10. CADENAS DE MÁRKOV

De acuerdo con Avilés (2017), el comportamiento de las cadenas de Márkov está gobernado por un conjunto de matrices de probabilidades de transición que indican las probabilidades de ocurrencia de los estados de un sistema para un intervalo de tiempo del futuro a partir de la información del estado actual y/o estados de los intervalos pasados, dependiendo de la orden del modelo. La propiedad Markoviana del modelo de cadena de Márkov de m^{th} orden es:

$$P(Y_{tn}/Y_{tn-1}, Y_{tn-2}, Y_{tn-3}, \dots, Y_1) = P(Y_{tn}/Y_{tn-1}, Y_{tn-2}, \dots, Y_{tn-m}) \quad (54)$$

Considerando una cadena de Márkov de primer orden (MCFO), es decir, $m=1$, las probabilidades de transición proporcionan el pronóstico probabilístico del estado un paso adelante basado en el estado actual, aplicando la siguiente fórmula:

$$p_{ij} = P(Y_{tn}/Y_{tn-1} = i) \quad (55)$$

Donde p_{ij} representa la probabilidad de transición que Y_{tn} es igual a la categoría j dado que Y_{tn-1} es igual a la categoría i . La probabilidad de transición estimada \hat{p}_{ij} puede ser calculada tomando en cuenta las frecuencias relativas condicionales de las transiciones f_{ij} :

$$\hat{p}_{ij} = \frac{f_{ij}}{\sum_j f_{ij}} \quad i, j = 1, \dots, s \quad (56)$$

Donde f_{ij} es la frecuencia que Y es igual a la categoría i en el tiempo t_{n-1} e igual a la categoría j en el tiempo t_n . El valor de s es el número de estados del sistema. El numerador presenta el número de transiciones de categoría i a j y el denominador representa la suma del número de transiciones de la categoría i a cualquier otra categoría.

III. MATERIALES Y MÉTODOS

3.1. UBICACIÓN DE LA ZONA DE ESTUDIO

La Cuenca Hidrográfica del Rio Chancay- Lambayeque se encuentra localizada en el Norte del territorio Peruano, en la región Nororiental. Políticamente, está ubicada en los departamentos de Lambayeque, provincias de Chiclayo, Lambayeque y Ferreñafe, y Cajamarca, provincias de Chota, Santa Cruz, Hualgayoc y San Miguel.

Geográficamente, se encuentra entre los paralelos 6° 20' y 6° 56' de Latitud Sur, y meridianos 78° 38' y 80° 00' de Longitud Oeste, extendiéndose entre el nivel del mar y los 3950 msnm.

Hidrográficamente, pertenece a la Vertiente del Pacífico, pero desde el año de 1958 recibe las aportaciones del Rio Chontano y posteriormente del Rio Conchano, que pertenecen a la vertiente del Atlántico.

Limita al norte con la Cuenca del Rio La Leche, por el Sur con la Cuenca del Rio Jetepeque-Zaña, por el Este con la Cuenca del Chontano y por el Oeste con el Océano Pacífico.

La Cuenca del Rio Chancay- Lambayeque comprende a su vez el ámbito jurisdiccional del Distrito de Riego Chancay – Lambayeque el cual se subdivide en dos Sub Distritos de Riego: Regulado, que se encuentra en la parte baja del valle, y No regulado, que está comprendido en la parte alta del Valle.

3.2. DESCRIPCIÓN GENERAL DE LA CUENCA INTEGRADA CHANCAY-LAMBAYEQUE

La cuenca integrada del Rio Chancay – Lambayeque, está conformada por la cuenca natural del Rio Chancay (vertiente del Pacífico) y parte de la cuenca alta de los Ríos Chontano y Conchano (Vertiente del Atlántico) de quienes recibe mediante obra trasvase su aporte hídrico. Esto es conocido como el Sistema Chancay - Lambayeque, mediante el cual afianza hídricamente al Valle Chancay- Lambayeque, beneficiando la Irrigación del Sistema Tinajones.

La cuenca tiene como ámbito socioeconómico principal al valle Chancay- Lambayeque, en el cual destaca por su importancia la ciudad de Chiclayo. En la Costa las principales ciudades desarrolladas son Chiclayo, Lambayeque y Ferreñafe y en la parte alta las provincias de Chota y Santa Cruz. En el Sistema Chancay-Lambayeque se diferencian tres zonas, que se describen a continuación:

a. Zona Baja o Valle

Es una planicie con muy poca pendiente (0.1% a 2.0%) y se extiende desde el nivel del mar hasta los 500 msnm. Tiene un clima árido y seco con presencia casi nula de precipitaciones. En esta zona se encuentra el área irrigada del Proyecto Tinajones.

La zona muestra grandes contrastes, como áreas de riego o valle verde y otras desérticas o dunas, el bosque seco también es típico en esta zona. Los suelos son aluviales o eólicos, de perfiles variables en su textura con predominancia de suelos ligeros, arenosos, francos a franco limosos y arcillosos, lo que ha permitido el desarrollo agrícola del valle en base a cultivos de arroz, caña de azúcar, maíz, menestras, yuca, hortalizas, algodón, cucurbitáceas y frutales diversos; por el clima imperante es posible obtener dos cosechas al año.

La existencia de importantes restos arqueológicos comprueba que la agricultura bajo riego fue desarrollada desde las civilizaciones Pre-Incas. Actualmente, la agricultura en el valle sufre un proceso de modernización que ha comprometido la construcción de grandes obras de irrigación que constituyen el Sistema de Riego regulado, como el Reservorio Tinajones con capacidad de embalse de 320 MMC.

b. Zona Alta o Sierra

Comprende desde los 5000 hasta los 3500 msnm, su clima es de templado a frío con precipitaciones abundantes de Enero a Abril; durante el periodo de Junio a Setiembre las precipitaciones son casi nulas.

La topografía es muy accidentada, con pendientes fuertes. La agricultura en estas zonas es en secano, con cultivos de maíz, menestras, yuca, papa y olluco; en los valles templados se cultivan frutales diversos, como paltas, plátanos, naranjas, chirimoyas, caña de azúcar. La producción en esta zona es de autoconsumo; en algunas zonas existen pequeñas obras de riego.

c. Zona del Traslase

La zona del traslase alcanza un área de 393 Km², de las cuales 391Km², corresponde a la cuenca alta y media del Río Chontano y 2 Km² a la parte de la cuenca del Río Conchano; cuyos aportes hídricos se incorporan al Río Chancay por medio de dos (02) túneles, uno para el traslase de los aportes del Río Conchano al Chontano y el otro para el traslase del Chontano al Chancay.

La derivación del Río Conchano hacia el Río Chontano es a partir de la quebrada Salabamba, mediante un túnel que desemboca en el Río Doña Ana, que es el principal afluente de la Cuenca alta del Chontano.

Con respecto a la hidrografía, podemos decir que el Río Chancay, de régimen irregular, nace en la laguna Mishacocha, a 3900 msnm, ubicada en la Codillera Occidental de los Andes. Tiene una longitud de 205 Km; en su recorrido recibe los aportes de los Ríos Huambayaco, Cirato, Cumbil y Quebrada Chancay por la margen derecha; y los Ríos Cañad y San Lorenzo por la margen izquierda.

El Río Chontano tiene su origen en la confluencia de las Quebradas Río Grande y Yanayacu, su longitud es de 35 Km hasta su derivación al Río Chancay. Sus principales afluentes son: por la margen derecha la Quebrada Doña Ana, cauce donde desemboca el Túnel Conchano y por la margen izquierda, el Río Jalqueño.

El Río Conchano, descarga sus aguas en el Río Maraño, por medio del Río Silaco, tiene su origen en varios afloramientos de agua de procedencia subterránea, cuya extensión es de 2 Km².

3.3. DESCRIPCIÓN GENERAL DEL SISTEMA TINAJONES

Las agua superficiales del Río Chancay y sus derivaciones (Chontano y Conchano) son captadas en la Bocatoma Raca Rumi, en donde los excedentes de las aguas circulantes después de satisfacer la demanda, son derivadas por el Canal Alimentador hacia el Reservorio Tinajones (capacidad de 320 MMC) para ser almacenadas; luego, en la regulación las aguas son devueltas al Río Chancay, por el canal de descarga, para ser distribuidas al valle a través de la infraestructura de canales, mediante los Repartidores La Puntilla y Desaguadero. El Repartidor La Puntilla reparte agua para el Río Reque, que es la

continuación del Río Chancay; y al canal Taymi, que conduce las aguas hasta el Repartidor Desaguadero, de donde se distribuye las aguas a los canales Taymi, Lambayeque y Patapo. Estos canales se dividen y subdividen en otros para abarcar toda el área de riego.

a. Reservorio Tinajones

Se encuentra ubicado en la margen derecha del Río Chancay, ha sido construido en un área de 20 Km², con una capacidad de almacenamiento de 320 MMC, formándose el embalse por medio de un dique principal de estructura de tierra zonificada de 41 m de altura y 2.38 Km de longitud y tres diques secundarios. Las obras conexas son: aliviadero de demasías y canal evacuador, que permiten la evacuación de las aguas cuando el reservorio está a plena capacidad.

El Reservorio Tinajones y sus obras complementarias fueron construidas en el periodo 1963-1968 y están en servicio desde el año 1969. En más de 30 años de operación el Reservorio de Tinajones ha registrado volúmenes mínimos y volúmenes normales de agua embalsada.

b. Bocatoma Raca Rumi

Es la obra principal y está ubicada a la altura de la cota 280 msnm del río Chancay, su estructura permite la captación de las aguas del río Chancay de hasta 75 m³/s, hacia el Canal Alimentador de Tinajones. Esta estructura ha soportado los eventos extraordinarios de los Fenómenos El Niño 1983 y 1998. Fue construida en el periodo 1966-1968 y están en servicio desde el año 1969.

c. Canal de alimentación

Conduce las aguas derivadas del río Chancay hacia el reservorio Tinajones. Se inicia en la Bocatoma Raca Rumi, inmediatamente después de la estructura de salida del desarenador, termina en la estructura de entrega al Reservorio Tinajones, tiene una longitud de 16080 km y una capacidad máxima de 70 m³/s. No se cuenta con Manual de Operación y Mantenimiento. En general las obras del canal están en buen estado de conservación.

d. Canal de descarga

Conduce las aguas del reservorio hacia el Río Chancay para ser distribuido a todo el valle. Este canal tiene una longitud de 4,00 Km y una capacidad de descarga de 70 m³/s.

e. Obras de derivación

El túnel Chontano permite derivar las aguas del Chontano al río Chancay, tiene una longitud de 4.76 Km con una capacidad de conducción de $32 \text{ m}^3/\text{s}$, revestido de concreto simple; permite el trasvase de 200 MMC (incluyendo las aguas de trasvase del río Conchano de 90 MMC). Fue construido entre los años 1954 y 1960.

El túnel Conchano permite derivar las aguas del río Conchano (90 MMC) al cauce del Río Chontano, tiene una longitud de 4213 m, su capacidad máxima es de $13 \text{ m}^3/\text{s}$, la sección es circular de 2,5 m de diámetro y revestido de concreto, y obras conexas. Fue construida entre los años 1975 y 1982.

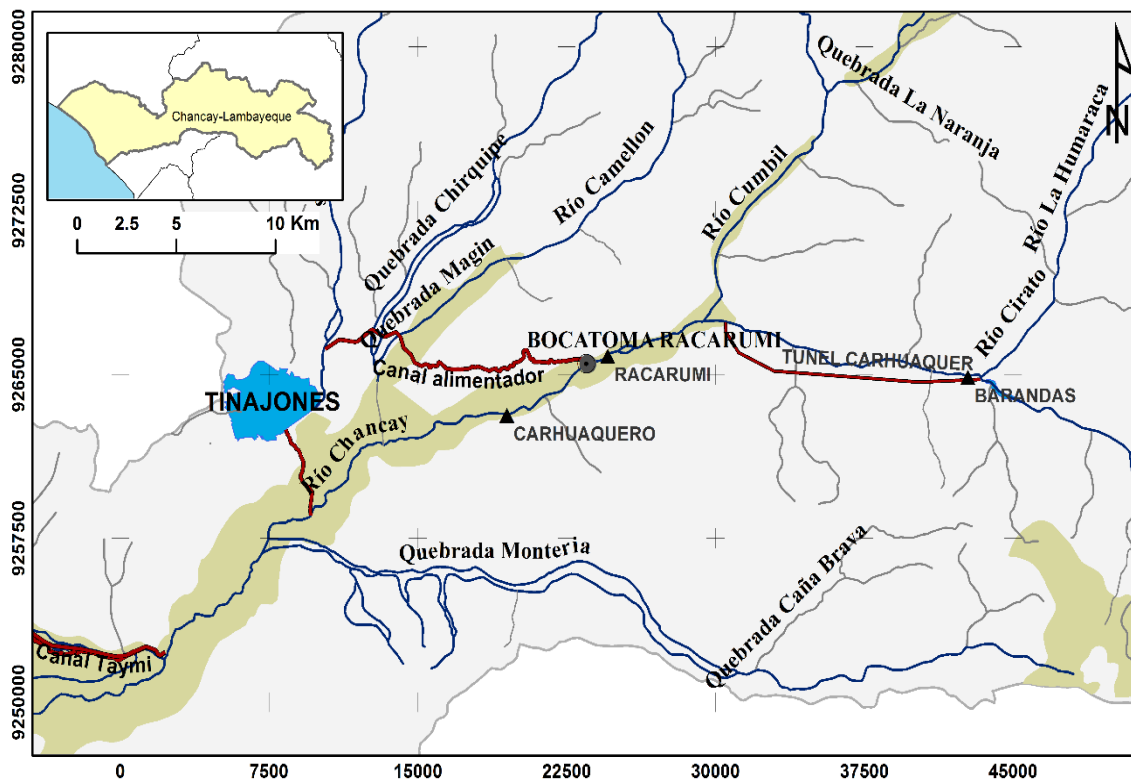


Figura 5. Zona de estudio – Cuenca Chancay Lambayeque.

3.4. HIDROMETRÍA DEL RÍO CHANCA Y-LAMBAYEQUE

La información de volúmenes anuales del río Chancay Lambayeque desde el 1914 hasta 2019 corresponden a la estación Raca Rumi, bocatoma del mismo nombre, y son monitorizadas por el Proyecto Especial Olmos Tinajones (Cuadro 3).

Cuadro 3. Volúmenes anuales registrados en la estación Raca Rumi.

Año	Volumen (mmc)	Año	Volumen (mmc)	Año	Volumen (mmc)
1914	999.648	1950	497.954	1986	1003.778
1915	1069.373	1951	544.303	1987	914.657
1916	774.645	1952	664.658	1988	928.464
1917	1374.538	1953	1042.394	1989	1332.018
1918	1051.384	1954	670.896	1990	718.685
1919	1096.502	1955	657.496	1991	769.652
1920	929.936	1956	831.006	1992	668.722
1921	1146.963	1957	747.953	1993	1438.145
1922	1229.947	1958	685.147	1994	1227.989
1923	855.325	1959	698.311	1995	809.806
1924	825.941	1960	710.172	1996	1098.119
1925	3260.252	1961	644.176	1997	607.744
1926	783.685	1962	823.558	1998	1503.266
1927	968.799	1963	476.965	1999	1426.934
1928	1070.637	1964	913.763	2000	1223.739
1929	1125.731	1965	820.167	2001	1368.059
1930	1105.263	1966	663.531	2002	1223.414
1931	717.769	1967	808.124	2003	825.047
1932	1042.665	1968	468.856	2004	737.578
1933	1000.559	1969	951.158	2005	894.427
1934	850.233	1970	1051.347	2006	1238.239
1935	726.059	1971	1632.916	2007	1102.870
1936	788.533	1972	1046.257	2008	1631.356
1937	493.922	1973	1309.485	2009	1473.199
1938	792.176	1974	1040.093	2010	963.073
1939	821.100	1975	1754.886	2011	930.038
1940	698.982	1976	913.407	2012	1643.994
1941	1046.357	1977	1034.522	2013	1312.479
1942	664.658	1978	575.100	2014	991.333
1943	826.104	1979	775.378	2015	1266.179
1944	763.806	1980	531.159	2016	853.620
1945	685.796	1981	879.807	2017	1424.192
1946	813.046	1982	946.596	2018	974.744
1947	968.457	1983	1302.095	2019	1304.723
1948	822.415	1984	1501.561		
1949	650.036	1985	442.839		

Fuente: Proyecto Especial Olmos Tinajones (2020).

3.5. METODOLOGÍA

La metodología utilizada comprende dos etapas, la primera es la selección del modelo de serie de tiempo que mejor ha representado el pronóstico de los volúmenes anuales de la estación Raca Rumi en sus correspondientes muestras a largo plazo.

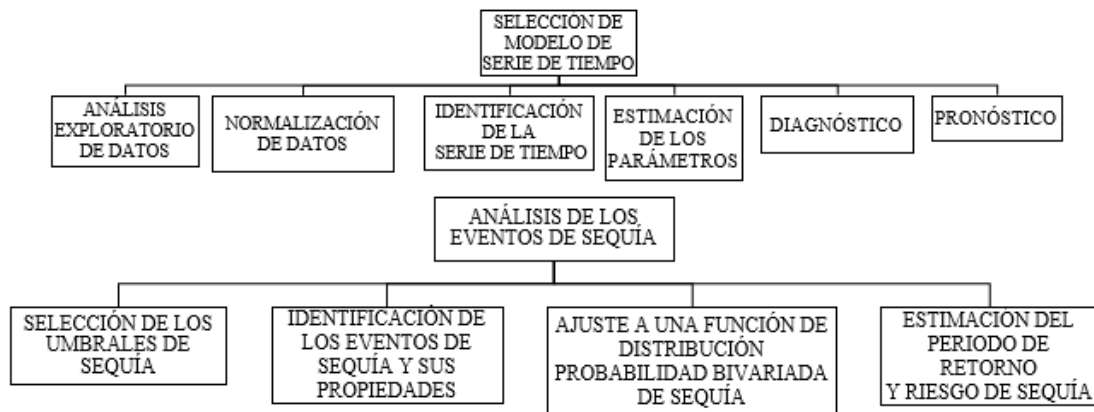


Figura 6. Esquema de la metodología para la selección del modelo de serie de tiempo y análisis de los eventos de sequía.

La segunda etapa de acuerdo con el método de riesgo de sequía propuesto por Salas *et al.* (2005), es el análisis de los eventos de sequía ocurridos para caracterizarlo según la duración y el déficit, medio o acumulado, respecto a un umbral (CEDEX 2017) posteriormente estimar de la probabilidad que ocurra un déficit D mayor a uno específico D_0 e intensidad I mayor a una específica I_0 así como su correspondiente periodo de retorno y riesgo de sequía.

3.5.1. Selección de modelo de serie de tiempo

a. Análisis exploratorio de datos

La finalidad del análisis exploratorio de la serie de tiempo es examinar los volúmenes anuales registrados en la estación Raca Rumi y, conseguir un entendimiento básico y las relaciones existentes entre las variables (Salvador y Gargallo 2003) previo a la aplicación de técnicas de modelamiento de series de tiempo.

La serie de volumen anual consiste en dos periodos, el primero desde 1914 hasta 1969 antes de entrar en funcionamiento el Sistema Tinajones, y el segundo desde 1970 hasta

2019, transcurso que viene operando. Primero se evaluó la consistencia en la media y la desviación estándar y el análisis de tendencia de ambos periodos mediante las pruebas estadísticas de T Student y F de Fisher y el coeficiente de correlación.

Dada que la media y desviación estándar de los periodos de estudio no son diferentes estadísticamente entonces no se hicieron las correcciones respectivas. Este mismo tratamiento se efectuó nuevamente como bondad de la información corregida en el caso de la transformación de la serie durante la etapa de normalización.

Luego se determinó las principales estadísticas de la muestra como la media aritmética, desviación estándar, varianza, coeficiente de variación y de asimetría y la función de autocorrelación. Así mismo, se obtuvo las estadísticas de sequía como son el periodo más seco y húmedo, máximo superávit y el déficit histórico.

Enseguida se representaron gráficamente por un diagrama de caja e histogramas el comportamiento de la serie de tiempo, para distinguir su posición central y dispersión como “datos atípicos”, tendencias y/o cambios hayan sucedido.

De otro lado, se examinó el grado de interrelación existente entre los volúmenes anuales como la distribución empírica de los datos observados que permita representar la función de distribución de probabilidad normal de la población origen de donde se asumen provienen.

b. Normalización de los datos

La normalización de los datos consistió en la transformación de escala de la distribución de una variable con el objetivo de poder hacer comparaciones de los valores normalizados con conjuntos de datos de manera que elimine los efectos de influencias.

Si la desviación estándar aparenta ser directamente proporcional a la media, es útil una transformación logarítmica, un modelo multiplicativo estabiliza la desviación estándar y de esa manera se alcanzará una varianza constante a través del tiempo. (Bastidas 2015) se dispuso de la siguiente transformación logarítmica:

$$Y = Ln(X + a) \tag{57}$$

Donde Y es la serie hidrológica normalizada, X la serie hidrológica original observada y a el coeficiente de transformación cuyo valor determino el mejor ajuste lineal para repetir y comprobar satisfactoriamente las pruebas de normalidad (Sveinsson *et al.* 2007).

Los datos observados y los transformados en el periodo 1969 a 2019 fueron sometidos a dos pruebas de normalidad, la prueba de correlación de la probabilidad de posición PPCC o prueba de Filliben y la prueba de asimetría a fin de conocer su inclinación o ajuste lineal de la PDF Normal.

La prueba de correlación de la probabilidad de posición usa el coeficiente de correlación r entre las observaciones ordenadas X_i , y los correspondientes cuantiles ajustados M_i determinados con el ploteo de posición p_i por cada X_i (Heo *et al.* 2008). Se supone que las observaciones podrían haber sido extraídos de la distribución adaptada (normal o lognormal) si el valor de r es cercano a 1. En esencia, r mide la linealidad de la gráfica de probabilidad, provee una evaluación cuantitativa del ajuste. El coeficiente de correlación r está dado por:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(M_i - \bar{M})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 (M_i - \bar{M})^2}} \quad (58)$$

Donde \bar{X} y \bar{M} denotan las medias de los valores observados X_i y los cuantiles ajustados M_i , respectivamente, y n es el tamaño de la muestra. En la ecuación para estimar el orden estadístico de la mediana por M_i se utiliza:

$$M_i = \Phi^{-1}(m_i) \quad (59)$$

Donde, Φ^{-1} es la función acumulada inversa de la distribución normal estándar, y m_i es el valor de la mediana propuesto por Filliben (1975) citado por Heo *et al.* (2008):

$$m_i = 1 - (0.5)^{\frac{1}{n}}, \quad i = 1 \quad (60)$$

$$m_i = \frac{(i - 0.375)}{(n + 0.365)}, \quad i = 2, n - 1 \quad (61)$$

$$m_i = (0.5)^{\frac{1}{n}}, \quad i = n \quad (62)$$

La hipótesis nula de una muestra dada no puede ser rechazada si el nivel de significancia q (0.05) es considerado:

$$r > r_q(n) \quad (63)$$

Donde $r_q(n)$ es el estadístico de correlación de probabilidad de posición (PPCC) según el tamaño de la muestra y un nivel de significancia q .

El contraste del coeficiente de asimetría y curtosis a los valores observados y transformados, prueba si es cierta la hipótesis que los x_i son normales, los estadísticos muestrales dados por:

$$Z_{asimetria} = \frac{asimetria}{\sqrt{\frac{6}{N}}}, \quad Z_{curtosis} = \frac{curtosis}{\sqrt{\frac{24}{N}}} \quad (64)$$

Ambos se distribuyen asimétricamente según una $N(0,1)$ a un nivel de significancia del 10 % (Salvador y Gargallo 2003). En las pruebas de asimetría de normalidad donde la muestra de los X_i se distribuye asintóticamente según una $N(\mu_y, \sigma_y^2)$. Luego la asimetría muestral estimada se distribuye asintóticamente como $N(0, 6/N)$.

La hipótesis nula $H_0: g = 0$ versus $H_a: g \neq 0$ es rechazada al nivel de significancia α si $|g| > Z_{1-\alpha/2} * \sqrt{6/N}$

Finalmente, para representar con los datos normalizados el modelo matemático que describe la composición de la serie hidrológica anual, se calculó la componente estocástica o residual Z_t sustrayendo la media y dividiendo por la desviación estándar.

$$Z_t = \frac{Y_t - \bar{Y}}{S_y} \quad (65)$$

Donde \bar{Y} y S_y es la media y la desviación estándar de la serie transformada a normal.

Este proceso estandarizado Z_t puede ser independiente o no, es linealmente dependiente si sus valores sucesivos están relacionados por cualquier modelo mático lineal (Yevjevich 1972).

c. Identificación de la serie de tiempo

La identificación consiste en inducir, a partir de los datos, utilizando la función de autocorrelación FAC y la función de autocorrelación parcial FACP muestrales qué modelo o modelos tentativos ARMA se adaptarían mejor a las características de la serie de tiempo (Pérez 2010).

El análisis del correlograma, es una función entre los coeficientes de correlación serial ρ_k como ordenadas y el retardo o desfase k como abscisas, que describe analíticamente a las secuencias de valores continuos (en el sentido matemático) o de valores discretos y es usado para determinar la dependencia entre los valores sucesivos de la serie y lo podemos definir como la razón de la covarianza poblacional y la raíz cuadrada del producto de dos varianzas poblacionales (Aliaga 1980).

$$\rho_k = \frac{COV[Z_t Z_{t+k}]}{[VARZ_t VARZ_{t+k}]^{\frac{1}{2}}} \quad (66)$$

Los valores de ρ_k son estimados en hidrología desde las muestras, por una aproximación de las series directas, tal que el estimado del coeficiente de autocorrelación es:

$$r_k = \frac{\frac{1}{n-k} \sum_{t=1}^{n-k} Z_t Z_{t+k} - \frac{1}{(n-k)^2} (\sum_{t=1}^{n-k} Z_t) (\sum_{t=1}^{n-k} Z_{t+k})}{\left[\frac{1}{n-k} \sum_{t=1}^{n-k} Z_t^2 - \frac{1}{(n-k)^2} (\sum_{t=1}^{n-k} Z_t)^2 \right]^{\frac{1}{2}} \left[\frac{1}{n-k} \sum_{t=1}^{n-k} Z_{t+k}^2 - \left(\frac{1}{n-k} \sum_{t=1}^{n-k} Z_{t+k} \right)^2 \right]^{\frac{1}{2}}} \quad (67)$$

Donde r_k es el coeficiente de correlación serial muestral de retardo k y n el tamaño de la serie, $t = 1, 2, \dots, n$, Z_t la componente estocástica, k el retardo o desfase considerado.

La ecuación (67) da sesgado pero eficiente estimado de ρ_k .

De esta manera cuando se mide una variable a lo largo del tiempo, las observaciones en diferentes periodos con frecuencia están relacionadas o correlacionadas según el

coeficiente de autocorrelación r_k , así si los datos presentan tendencia las observaciones sucesivas están altamente correlacionadas y es típico encontrar el primer coeficiente de autocorrelación cercano a uno; también el coeficiente de autocorrelación para el retraso de tiempo dos será grande pero menor que el primero, y así sucesivamente con los otros valores (Pérez 2012).

De lo contrario, si los datos son estacionarios, los valores de autocorrelación decaen rápidamente hacia cero; por lo general, a partir del segundo o tercer rezago el coeficiente de autocorrelación es estadísticamente igual a cero.

Si los datos son aleatorios, todos los valores de autocorrelación son estadísticamente iguales a cero, por lo tanto, los valores sucesivos de la serie de tiempo no están relacionados entre sí.

Para probar la independencia de Z_t o si un coeficiente de autocorrelación es estadísticamente diferente de cero, se realizó lo siguiente:

1° Se planteó la hipótesis:

$$H_0: \rho_k = 0 \quad (68)$$

$$H_a: \rho_k \neq 0 \quad (69)$$

Si se rechaza H_0 , la autocorrelación entre Z_t y Z_{t-k} es significativa y por lo tanto los datos son ruido blanco. Si no se rechaza H_0 , la autocorrelación de orden k es estadísticamente igual a cero.

2° Se calculó el correlograma para Z_t según la ecuación (67)

3° Se calculó los límites de confianza LC del correlograma para un nivel de confianza $\alpha = 0.05$ es:

$$C = \frac{-1.96 \pm \sqrt{N - k - 1}}{N - k - 1} \quad (70)$$

4° Se calculó los estadísticos para una distribución T-Student:

$$T = \frac{r_k}{desv(r_k)} \quad (71)$$

Se rechaza H_0 si $|r_k| > t_{\alpha/2, v}$ donde v es el número de observaciones, menos 1. Es decir, que se rechaza H_0 si $|r_k| > desv(r_k)t_{(\frac{\alpha}{2}, v)}$

Si se desea probar que los datos son aleatorios o ruido blanco, la desviación estándar r_k será:

$$desv(r_k) = \sqrt{\frac{1 + 2 \sum_{i=1}^{k-1} r_i^2}{T}} \quad (72)$$

Donde m es el número de autocorrelaciones a calcular y.

$$desv(r_k) = \frac{1}{\sqrt{T}} \quad (73)$$

5° Se evaluaron los siguientes criterios de decisión:

Si al menos el 95% de los valores del correlograma de Z_t caen dentro de los límites de confianza, entonces se acepta la hipótesis planteada; vale decir, Z_t es independiente en el tiempo.

Si más del 5% de los valores calculados para el correlograma de Z_t caen fuera de los límites de confianza, entonces se rechaza la hipótesis planteada, aceptándose que Z_t es dependiente en el tiempo.

En este caso es necesario representar la dependencia de la componente estocástica Z_t , para lo cual se utilizan los modelos autorregresivos o markoviana de un orden adecuado.

Si Z_t resulta ser independiente, entonces ya no es necesario calcular los modelos de dependencia a lo que se denomina Márkov cero.

También se realizó una inspección visual para identificar los modelos AR (q) o MA (q) como indica (Pankratz 1983 citado por Alvarado 2007):

1° Los procesos estacionarios auto regresivos AR tienen FAC teóricas que cortan repentinamente el cero después de un número corto de lags o retrasos y FACP teóricas que cortan repentinamente el cero después de unos pocos retrasos y enseguida los valores de los lags se alternan en signo en lags posteriores. La longitud del último valor de la FAC antes de que la FACP corte el cero es al orden (p) del proceso AR.

2° El proceso de promedios móviles MA tiene una FAC que corta a cero después de un cierto número de lags. La longitud del último valor de la FAC antes de que la función corte a cero es el orden de (q) del proceso MA. Su FACP teórica cae más allá del cero y posteriormente sus valores se alternan en signo.

Consecuentemente, la memoria del sistema es un indicador de cuán grande es la dependencia del flujo observado actualmente con el pasado. Cuando los valores de la variable aleatoria es una función de su propio pasado, se dice que la variable exhibe dependencia serial (Goldman 1985). Idealmente, si el correlograma se aproxima a cero como un decaimiento suave de la función, se dice que tiene memoria larga, caso contrario un decaimiento brusco de la función el sistema tiene memoria corta.

d. Estimación de los parámetros

Los estimadores por momentos $\hat{\theta}_p$ y $\hat{\theta}_q$, $\hat{\mu}$ y $\hat{\sigma}_e^2$ de los parámetros de los modelos AR (1) y MA(1) de primer orden se estimaron según las ecuaciones (16), (24), (14) y (15 y 25) de la sección 2.6.3., sujeta a las condiciones de estacionalidad $|\theta_p| < 1$, e invertibilidad $|\theta_q| < 1$.

e. Diagnóstico

Establecido los modelos de series de tiempo, auto regresivo AR (1) y media móvil MA (1) se generó diez series de cincuenta años es decir de igual periodo de registro de la serie histórica, para verificar si el modelo seleccionado es capaz de representar la evolución de los volúmenes observados en el tiempo.

A fin de evaluar el término de error de un modelo estocástico ARMA es un proceso normal puramente aleatorio, es decir que su varianza es constante, su media es nula y no presenta autocorrelación ni correlación serial con la propia serie rezagada se debe cumplir dos condiciones:

- Los residuales del modelo deben ser ruido blanco o aleatorio.
- Los errores deben seguir una distribución normal.

La medida de evaluación aplicada a cada modelo de pronóstico de la variable se hizo con base a los indicadores del error cuadrático medio (MSE), desviación absoluta media (MAD) y el error porcentual absoluto medio (MAPE), estos errores de pronóstico se definen como la diferencia entre el valor observado Y_t y el valor estimado \hat{y} (Pérez 2010).

$$e_t = Y_t - \hat{Y}_t \quad (74)$$

- Error cuadrático medio (MSE) se define como:

$$\text{MSE} = \frac{1}{N} \sum_{t=1}^N e_t^2 \quad (75)$$

Donde N es el número de errores disponibles. El MSE penaliza los errores grandes.

- Desviación absoluta media (MAD) se define como:

$$\text{MAD} = \frac{1}{N} \sum_{t=1}^N |e_t| \quad (76)$$

El resultado obtenido se puede interpretar como el número de unidades promedio en las cuales cada pronóstico se desvía del valor real.

- Error porcentual absoluto medio (MAPE) se define como:

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^N \frac{|e_t|}{Y_t} * 100 \quad (77)$$

Entre las diferentes técnicas de pronóstico se eligió el que tuvo menor MSE, MAD y MAPE.

Asimismo, para identificar si los errores cumplen la condición de normalidad se realizó el contraste de hipótesis de Jarque Bera:

H_0 : Los errores siguen una distribución normal.

H_a : Los errores no siguen una distribución normal.

Se rechaza la hipótesis si el estadístico $JB > X^2_{\alpha}$. El estadístico de prueba es:

$$JB = \frac{N}{6} \left(A^2 + \frac{(k-3)^2}{4} \right) \quad (78)$$

Donde, A es el coeficiente de asimetría, K: es el valor de curtosis, N: es el número de datos.

También se verifico si los errores son ruido blanco calculando el estadístico Q de Ljung-Box (Pérez 2010) a los coeficientes de autocorrelación:

$$Q = N(N+2) \sum_{k=1}^m \frac{r_k^2}{N-k} \quad (79)$$

Donde m es el número de coeficientes de autocorrelación, k es el número de rezagos y r_k es el coeficiente de autocorrelación de orden k de los residuos.

El estadístico Q se distribuye asintóticamente como X^2 con (m-p-q) grados de libertad. El contraste de hipótesis asociado al estadístico Q es:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0 \quad (80)$$

$$H_a: \text{al menos un } \rho_k \neq 0, K = 1, 2, \dots, m. \quad (81)$$

Se rechaza la H_0 si $Q_m > X^2_{\alpha}$. Si no se rechaza la hipótesis nula, indica que no existe autocorrelación en los residuos hasta el orden m o al menos un coeficiente de autocorrelación es estadísticamente diferente de cero. No rechazar H_0 indica que las primeras m autocorrelaciones son estadísticamente iguales a cero, y, por lo tanto, los datos son ruido blanco.

Además, si el modelo cumple con asunciones referidas a las características implícitas de los residuales tales como la normalidad e independencia y es capaz de reproducir las

propiedades estadísticas históricas de los datos observados (Pérez 2010), se utilizaron los criterios de información Akaike (AIC) o el criterio de información de Akaike modificado (AICC) y el criterio información Schwartz (SIC) para comparar las asunciones y elegir el modelo que mantiene menor valor de ambos criterios.

– El AICC se define como:

$$AICC = n \cdot \ln \hat{\sigma}^2(\varepsilon) + \frac{2(k+1)n}{n-k-2} \quad (82)$$

Donde n es el tamaño de la muestra usada para el ajuste del modelo, k es el número de parámetros excluyendo el termino constante, $k = p+q$ para modelos ARMA (p,q) y $\hat{\sigma}^2(\varepsilon)$ es la varianza residual (sesgada) estimada por máxima verosimilitud. El AICC es eficiente pero consistente y bueno para muestras pequeñas, tiende a sobreajuste para k largos.

– Y el SIC como:

$$SIC = n \cdot \ln \hat{\sigma}^2(\varepsilon) + n + k \cdot \ln(n) \quad (83)$$

Donde n, k y $\hat{\sigma}^2(\varepsilon)$ son definidos de la misma forma para el estadístico AICC. En general el SIC es bueno para muestras grandes, pero tiende a “bajo ajuste” para muestras pequeñas. La eficiencia es usualmente más importante que la consistencia desde que el orden del modelo verdadero no es conocido por datos reales.

f. Pronostico

Una vez el modelo autorregresivo de primer orden ha sido estimado y validado hasta el periodo de 50 años, igual a la longitud de registro como de vida útil de la obra, se realizó la generación de series sintéticas anuales a largo plazo de 1500 años con 10 repeticiones basadas en los modelos de aproximaciones y esquemas.

La generación de volúmenes anuales con el modelo ARMA (p,q) tiene la siguiente estructura:

$$U_{v,t} = \sum_{j=1}^p \phi_{j,\tau} Z_{v,t-1} - \sum_{i=1}^q \theta_{i\tau} \sigma_{\varepsilon,\tau}^2 \varepsilon_{t-i} + \sigma_{\varepsilon\tau}^2 \varepsilon_{v,t} \quad (84)$$

En este caso, se aplicó una transformación inversa a la estandarización de los volúmenes anuales de agua.

$$Z_{v,\tau} = \frac{U_{v,\tau} - \bar{U}_\tau}{S(U)} \quad (85)$$

De manera que los volúmenes generados se obtuvieron de la siguiente forma:

$$Y_{v,\tau} = \bar{Y}_{v,\tau} + S_\tau(Y)Z_{v,\tau} \quad (86)$$

Y dependen de la media y desviación estándar histórica de los datos transformados y el parámetro estandarizado.

Si una transformación ha sido aplicada como la logarítmica, una transformación inversa se aplica y a partir de esta se obtiene los volúmenes anuales sintéticos.

$$X_{v,\tau} = \exp(Y_{v,\tau}) - a_\tau \quad (87)$$

3.5.2. Análisis de los eventos de sequía

El estudio de los eventos de sequía o déficits observados en la estación Raca Rumi consistió en evaluar su probabilidad de ocurrencia utilizando los déficits muestreados luego de truncar el nivel de la serie sintética de volúmenes anuales con los coeficientes de déficit λ asignados. El método que se sigue fue:

a. Selección de los niveles de truncamiento

Para determinar las anomalías o déficits de agua acontecidos d_j de la serie observada y las series sintética de 1500 años de muestra de los volúmenes de agua anual en la estación Raca Rumi, primero se establecieron los coeficientes de déficit λ iguales a 0.85, 0.95, 1.00 y 1.15 por ciento del volumen medio de agua anual \bar{X} .

$$X_0 = \lambda \bar{X} \quad (88)$$

La selección del nivel de truncamiento como la demanda del por ciento del volumen medio de agua anual es importante ya que define la sucesión de rachas o “runs” negativos que cambia la longitud y magnitud de la sequía crítica cuando cambia este nivel (Aliaga 1980).

b. Separación de rachas o runs

Con la comparación de los volúmenes anuales de agua disponible X_t y la demanda de agua X_0 por cada nivel de truncamiento, se obtuvo la sucesión de rachas o “runs”, positivo y negativo que determinan la existencia o no de años de sequía y, solo aquellas sucesiones negativas determinaron la anomalía del déficit de agua.

$$d_j = X_t - X_0 \quad (89)$$

c. Muestreo del déficit de agua (sequía)

De la sucesión de rachas o “runs” negativos o anomalías de agua d_j por nivel de truncamiento se muestreo los déficits acumulados de agua D_0 e intensidades I_0 de duración lo igual a 1,2,3 y 4 años consecutivos.

De esta manera según los coeficientes de déficit λ asignados se obtuvieron los valores de déficit de agua observado (histórica) y sintéticos (generados) a largo plazo. La primera se utilizó para definir los valores de déficit según su percentil a fin de evaluar su probabilidad de ocurrencia posteriormente y, la segunda se utilizó para el análisis de los eventos de sequía a largo plazo.

d. Transición de sequías

A partir de la sucesión de runs positivos y negativos, se obtuvieron las secuencias de superávit (denotado por 1) y déficit (denotado por 0) para contar el número de transiciones N de un estado de falla (f) a uno de seguro (s) o 0 a 1, esto es el número de veces de transiciones que ocurre de un evento de sequía a un evento de superávit. Además de este proceso se contó las transiciones de un estado de falla (f) a otro de falla (f) y de uno seguro (s) a otro seguro (s).

Enseguida se calculó las probabilidades de transición por cada serie de déficit como:

$$p_{fs} = p_{01} = \frac{N_{fs}}{N_{ff} + N_{fs}} \quad (90)$$

$$p_{sf} = p_{10} = \frac{N_{sf}}{N_{ss} + N_{sf}} \quad (91)$$

Posteriormente la ocurrencia de una sequía de determinada duración (variable discreta) se representa con una distribución geométrica, ajustando cada probabilidad de transición de año húmedo siendo el anterior seco p_{01} o de año seco siendo el anterior húmedo p_{10} donde el parámetro de cada periodo es 0.50.

e. Ajuste a una distribución de probabilidad bivariado de sequías

Las sequías o déficit acumulado como las intensidades generadas agrupadas en cuatro categorías según las duraciones en años $l_o = 1, 2, 3$ y 4 años, fueron sometidas a la prueba de Grubbs para determinar valores atípicos, enseguida se ajustó a la función de distribución de probabilidad Gamma, los parámetros de forma γ y escala β de la distribución se obtuvieron aplicando el método de máxima verosimilitud comprobando su bondad de ajuste con la prueba de Anderson Darling con un nivel de significancia del 5%

f. Estimación de la probabilidad de excedencia de los eventos de sequía acontecidos

Con los percentiles de déficit D_o e intensidad I_o observados de duración $l_o=1,2,3$ y 4 años se obtuvieron la probabilidad de excedencia de una sequía en función de su duración y déficit dada por la integral del producto de las funciones de densidad como se expresa en la ecuación 89 y 90.

$$P(D > D_o, L = l_o) = \int_{D_o}^{\infty} \frac{1}{\beta \Gamma(l_o \Gamma)} \left(\frac{z}{\beta}\right)^{l_o \Gamma - 1} e^{-z/\beta} p_{01} (1 - p_{01})^{l_o - 1} dz \quad (92)$$

$$P(I > I_o, L = l_o) = \int_{I_o}^{\infty} \frac{1}{\beta \Gamma(l_o \Gamma)} \left(\frac{l_o z}{\beta}\right)^{l_o \Gamma - 1} e^{-l_o z/\beta} p_{01} (1 - p_{01})^{l_o - 1} dz \quad (93)$$

g. Periodo de retorno y riesgo de sequía

Por último, se calculó el periodo de retorno T (Salas *et al.* 2005) y el riesgo R en n años (Chow *et al.* 1988) de un evento de déficit e intensidad de sequía dada:

$$T = \frac{p_{01} + p_{10}}{p_{01}p_{10}} \frac{1}{P(E)} = \frac{p_{01} + p_{10}}{p_{01}p_{10}} \frac{1}{P(D > D_o, L = l_o)} \quad (94)$$

$$T = \frac{p_{01} + p_{10}}{p_{01}p_{10}} \frac{1}{P(E)} = \frac{p_{01} + p_{10}}{p_{01}p_{10}} \frac{1}{P(I > I_o, L = l_o)} \quad (95)$$

$$R = 1 - \left(1 - \frac{1}{T}\right)^n \quad (96)$$

IV. RESULTADOS Y DISCUSIÓN

4.1. SELECCIÓN DE MODELO DE SERIE DE TIEMPO

a. Análisis exploratorio de datos

Para la selección del modelo de serie de tiempo, promedios móviles $MA(q)$, se contó con los volúmenes anuales de la estación Raca Rumi, registrados desde el año 1914 a 2019 (Figura 7).

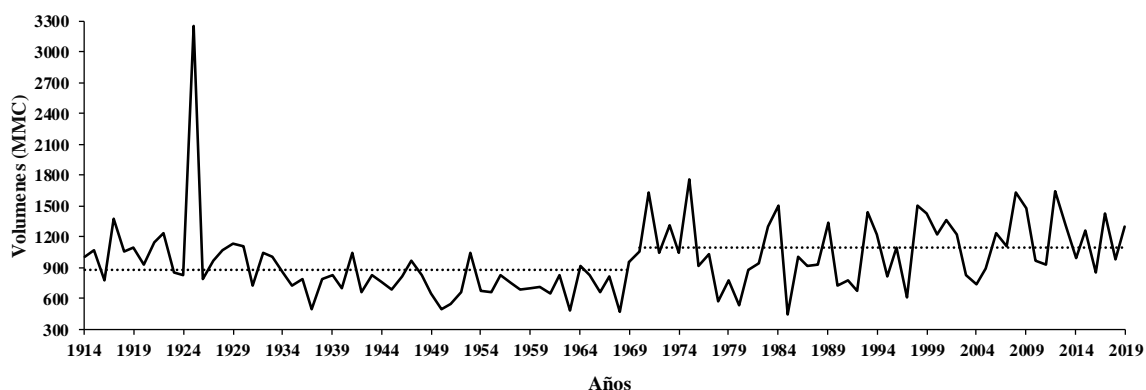


Figura 7. Hidrograma histórico registrado en la Estación Raca Rumi, Lambayeque.

Las oscilaciones de los volúmenes de agua en un nivel constante durante el tiempo de registro, se trataría de una serie univariada y estacionaria (Peña 2010), con incertidumbre de las variaciones temporales del escurrimiento (Sveisson *et al.* 2007) precedente de los suministros de agua con sucesos de depresión o escasez para su uso.

Asimismo, se observa dos periodos en la serie, el primero de 1914 a 1969 antes de la operación del embalse con valores atípicos como del año 1925, a consecuencia del fenómeno del niño. El segundo de 1970 a 2019 corresponde al inicio de la operación del sistema hidráulico Chancay – Lambayeque (PEOT 2011). con un salto por el incremento de volumen de agua del trasvase de los ríos Chontano y Conchano.

La consistencia en la media y la desviación estándar de los dos periodos en la serie observada ha demostrado que las medias han sido distintas y la desviación estándar no lo son, es decir su composición y estructura no son uniformes.

Cuadro 4. Prueba en la media y desviación estándar primer y segundo periodo de la serie.

Consistencia	Prueba estadística		
	F_c	F_t	Significativo
Desviación Estándar	1.422	1.592	Sí
Media	T_c	T_t	Significativo
	3.121	1.660	No

En consecuencia, las medias diferentes de los periodos conocidos pueden considerarse como una detección de un cambio (Maidment *et al.* 1993), mientras la serie temporal de operación del sistema hidráulico 1970 a 2019, la media y desviación estándar son iguales, es decir los volúmenes de agua registrados serían homogéneos y provienen de la misma población (Aliaga 1985).

Cuadro 5. Prueba en la media y desviación estándar segundo periodo de la serie.

Consistencia	Prueba estadística		
	F_c	F_t	Significativo
Media	1.387	1.677	Si
Desviación Estándar	T_c	T_t	Significativo
	1.416	1.984	Si

De la misma forma, la tendencia para el primer periodo ha sido no significativo, sin embargo, para el segundo periodo los datos mantienen una correspondencia recíproca con el tiempo.

Cuadro 6. Resumen de la prueba en el coeficiente de correlación.

Periodo	T_c	T_t	Prueba
1914-2019	-1.946	1.677	R no es significativo
1970-2019	46.932	1.677	R es significativo

La serie exhibe una tendencia lineal en la media y la desviación estándar, de acuerdo con los estadísticos calculados y tabulados con un 95% de probabilidad, el coeficiente de correlación R resulta significativo siendo necesario su corrección y estandarización.

Cuadro 7. Estadísticos de la serie histórica y transformada de los volúmenes del río Chancay-Lambayeque (MMC) periodo 1970 -2019.

Estadísticas	Serie	Serie
	Histórica	Transformada
Media	1091.35	8.26
Desviación estándar	314.60	0.08
Coefficiente de variación	0.29	0.01
Coefficiente de asimetría	0.09	-0.06
Coefficiente de autocorrelación	0.068	0.075
Sequía más larga*	7.00	7.00
Máximo déficit*	1983.51	0.52

*Nivel de demanda igual a la media

En el Cuadro 7 se muestran los resultados del análisis estadístico de los volúmenes anuales del río Chancay Lambayeque, la media ha sido 1091.35 MMC con una desviación estándar 314.60 MMC y un coeficiente de variación de 0.29, se atribuye la media es representativa de los datos y son homogéneos de una supuesta asimetría simétrica (Figura 8).

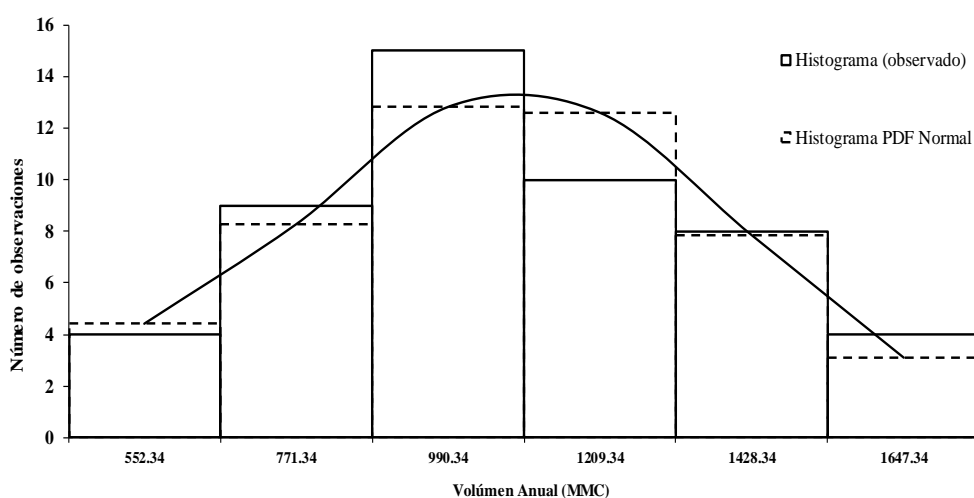


Figura 8. Histograma del volumen anual MMC de la estación Raca Rumi.

El coeficiente de autocorrelación de primer orden ha sido 0.068, cercano o fluctuante alrededor de cero, pero no necesariamente igual, debido a la variabilidad muestral pertenecería a una serie independiente, no afectada por el origen del momento de los datos (Salvador y Gargallo 2003).

En efecto, Salas (2000) sostiene el coeficiente de correlación de retardo 1 da una medida del grado de persistencia y son valores positivos y pequeños para los volúmenes anuales de los ríos de zonas áridas y semiáridas, incluso pueden presentar valores diversos, un valor promedio de 0.03, las secuencias de descargas anuales son aleatorias sin persistencia debido a la inexistencia o al efecto insignificante del agua subterránea, sin embargo en ríos de otras zonas indican lo contrario con un valor promedio de 0.15, en consecuencia, esta medida dependerá del tamaño de la cuenca, factores geológicos y su posición respecto a almacenamientos de agua superficial y subterránea o la presencia de embalses.

b. Normalización de datos

La normalización de los datos consistió en comprobar la normalidad $N(0,1)$ de la distribución de los volúmenes anuales Figura 5, según el contraste de asimetría y curtosis a un nivel de significancia 10%, los estadísticos muestrales demostraron no se distribuyen asintóticamente, son sesgados (Cuadro 8).

Cuadro 8. Prueba de asimetría y curtosis.

Serie	Z_a	Z_c	Prueba
Histórica	1.67	0.74	Se rechaza la H_0
Transformada	0.27	0.74	Se acepta la H_0

Sin embargo, la falta de normalidad de la serie histórica fue corregida mediante una transformación logarítmica de la forma $Y_t = \log X_t + 2793.0615$, comprobando su mejor ajuste a la posición gráfica de los cuantiles de la serie hidrológica transformada Figura 9 con el coeficiente de correlación mediante la prueba de Filliben (Cuadro 8).

El uso de las transformaciones como cambio de forma para la distribución de los datos es útil en la determinación empírica de la linealidad de los datos originales como una distribución normal y homocedasticidad (Sakia 1992).

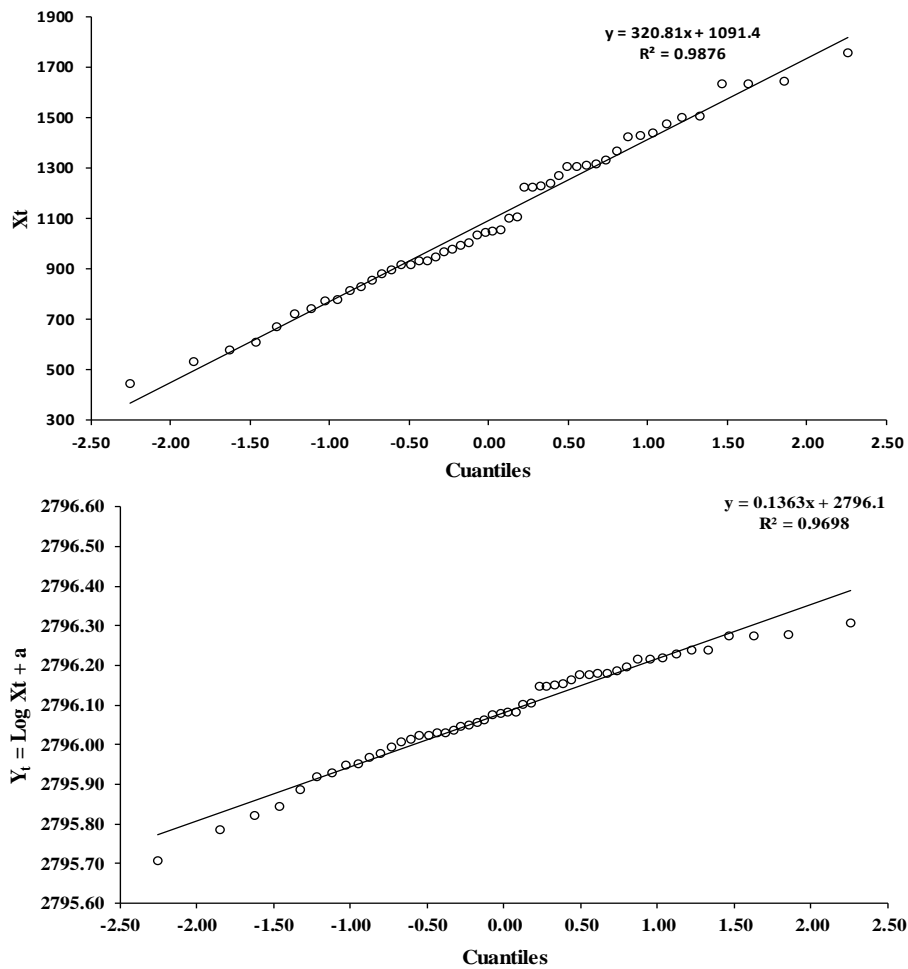


Figura 9. Diagrama de cuantiles de la serie histórica y transformada

Es conveniente notar que la transformación realizada esta reconocida para datos de asimetría positiva y la elección de la frecuencia empírica de posición elegida mejor influyó en el ploteo de probabilidad (Castillo y Lozano 2007).

Cuadro 9. Prueba de Filliben en la serie de histórica y transformada.

Serie	r	r _q	Prueba
Histórica	0.994	0.980	Se acepta la Ho
Transformada	0.985	0.980	Se acepta la Ho

Así, los estadísticos descriptivos de la serie transformada presentaron un coeficiente de variación de 0.01, homogéneo de asimetría negativa, es decir hay valores separados de la media a la izquierda con una desviación cercana a cero, la mayor parte de los datos se agrupan cercanos a su media, Cuadro 6. La gráfica de cajas no muestra valores atípicos

fuera de los límites o bigotes, el tamaño de la caja nos indica la alta variabilidad de los volúmenes anuales por encima y debajo del rango intercuartil (Figura 10).

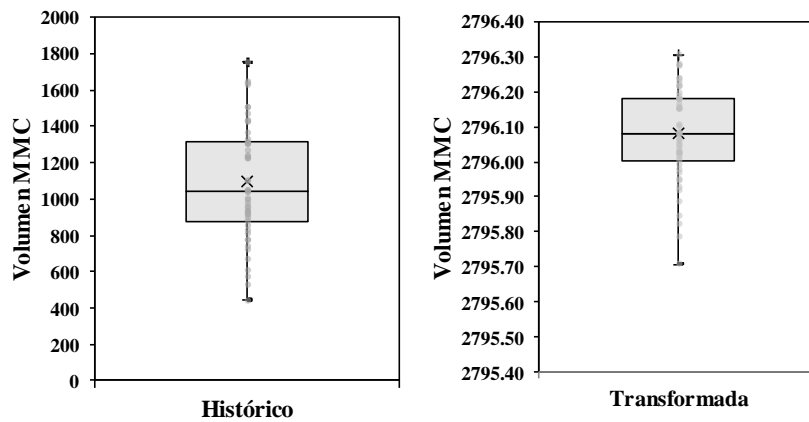


Figura 10. Diagrama de caja serie histórico y transformada.

Para determinar el modelo de series de tiempo univariado AR (1) como MA (1) y la estimación de sus parámetros, se obtuvo la componente estocástica Z_t , estandarizando la serie transformada, restando el promedio y dividiendo por la desviación estándar.

c. Identificación de la serie de tiempo

Para identificar el modelo de serie de tiempo se obtuvo la función de autocorrelación, como se observa en la Figura 11, los valores de los primeros retrasos caen progresivamente, cortando el cero luego del segundo retraso después cambian en signos en diferentes retrasos, este comportamiento del proceso lo define de corta memoria típico de un proceso estacionario.

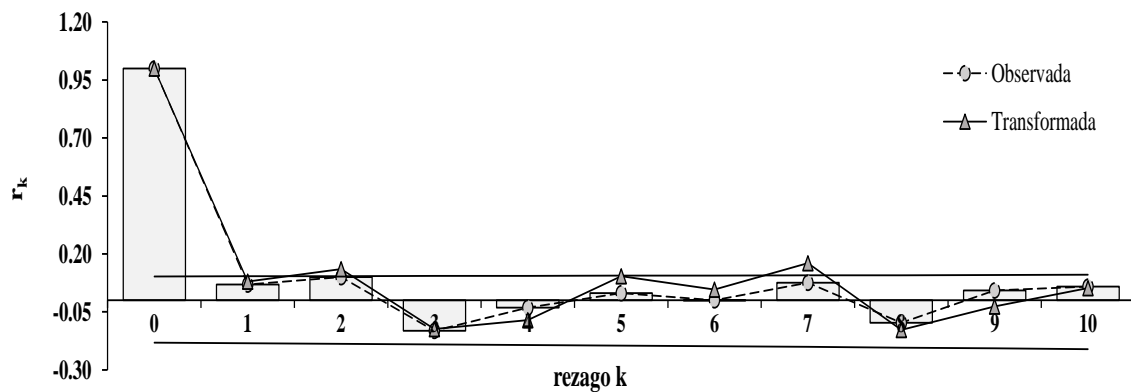


Figura 11. Función de autocorrelación FAC de la serie observada y transformada.

Cuadro 10. Prueba de significancia de la función de autocorrelación.

k	r_k	t.desv(r_k)	Prueba
0	1.000	-0.284	
1	0.068	-0.286	
2	0.101	-0.288	
3	-0.131	-0.293	
4	-0.032	-0.293	
5	0.031	-0.294	fac es significativo,
6	0.000	-0.294	los datos no son ruido blanco
7	0.076	-0.295	
8	-0.096	-0.298	
9	0.043	-0.298	
10	0.059	-0.299	

Para probar cuales son los valores r_k de autocorrelación significativos que superen el producto t.desv(r_k) se realizó el contraste de significancia Cuadro 9, así se tiene la autocorrelación entre los datos son significativos no son ruido blanco (Pérez 2012).

d. Estimación de los parámetros

En el Cuadro 11 se muestran los parámetros ϕ y θ determinados por el método de momentos, para los modelos de análisis de series de tiempo AR (1) y MA (1) de media sustraída y estandarizada.

Cuadro 11. Parámetros de los modelos univariados autorregresivos y media móviles.

Parámetros	AR (1)	MA (1)
Media	8.2614	8.2614
Varianza	0.0066	0.0066
Ruido blanco	0.9944	0.9944
phi	0.0749	-
theta	-	-0.0753

Además, se muestra la media y varianza del ajuste a la función de distribución de probabilidad Log Normal como la varianza y ruido blanco de los errores probados de los

modelos univariados adoptados. Estos parámetros se reemplazaron en las ecuaciones de los modelos matemáticos (Cuadro 12).

Cuadro 12. Ecuación de los modelos univariados.

Modelo	Ecuación
AR (1)	$Y_t = 8.2614 + 0.0749(Y_{t-1} - 8.2614) + \varepsilon_t$
MA (1)	$Y_t = 8.2614 - 0.0753(Y_{t-1} - 8.2614) + \varepsilon_t$

Las series fueron generadas de la ecuación de cada modelo e invertidas al dominio de volumen original por $X_t = 10^{(Y_t - 2793.0615)}$

e. Diagnóstico

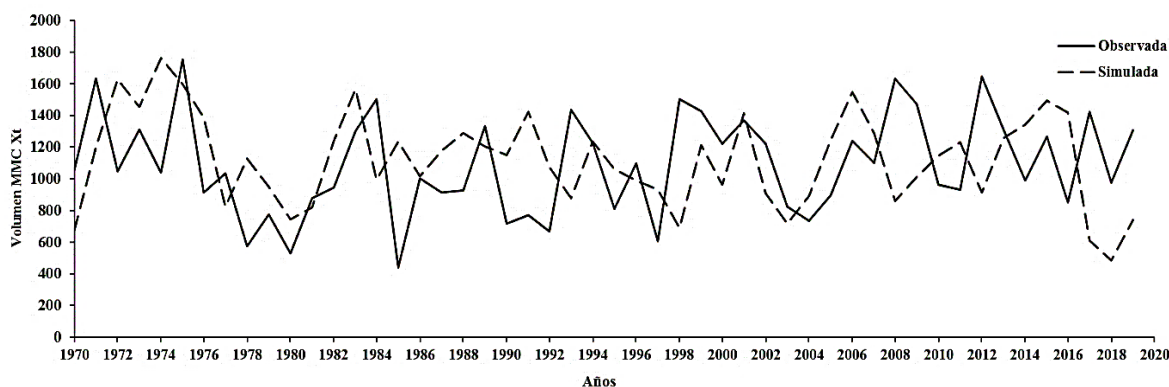


Figura 12. Comparación serie simulada vs serie observada

Se analizaron los residuos obtenidos de la comparación entre los valores de volúmenes anuales calculados del modelo y los valores observados de la serie histórica Figura 12, sin tener en cuenta la estructura estadística de la población como la bondad de las predicciones (Pérez 2012). El menor error de la predicción en las tres técnicas de medida ha sido con el modelo MA (1). Tanto en la medición del tamaño, la dispersión y la desviación en términos porcentuales del error de pronóstico.

Cuadro 13. Medidas de evaluación de los modelos univariados.

Modelo	MSE	MAD	MAPE
AR (1)	209808.13	373.02	37.72
MA (1)	175254.27	354.66	36.21

Además, se ha contrastado los residuales del modelo adoptado deben ser aleatorios y los errores siguen una distribución normal según la prueba de Jarque Bera (Cuadro 14).

Cuadro 14. Prueba de Jarque Bera a las técnicas de pronóstico seleccionados.

Modelo	A	K	JB	X ² _c	Prueba
AR (1)	0.172	2.967	0.247	5.991	Los errores siguen PDF Normal
MA (1)	0.112	2.377	0.913	5.991	Los errores siguen PDF Normal

Finalmente, de acuerdo con los criterios AICC y SIC, se eligió el modelo con el criterio que mantiene menor valor, en este estudio ambos modelos AR (1) y MA (1) cumplen la condición si el modelo asumió las características propias de los residuales en normalidad e independencia como la capacidad de reproducir las estadísticas históricas.

Cuadro 15. Criterio de información de Akaike (AIC) y Schwarz (SBC).

Modelo	AICC	SIC
AR (1)	53.97	53.63
MA (1)	53.97	53.63

Por tanto, el modelo adoptado que mejor representó la serie temporal de volúmenes anuales es la media móvil MA (1) luego que satisface las condiciones anteriores, este se utilizó como instrumento de pronóstico.

f. Pronóstico

El pronóstico ha consistido en predecir los volúmenes anuales de la estación Raca Rumi adoptando el modelo media móvil MA (1). Se han generado series sintéticas de 1500 años con diez repeticiones sujeta al menor error posible, para el cálculo de la probabilidad de ocurrencia de la sequía de un déficit y duración determinada.

Las estadísticas de la serie sintética son similares a la serie observada, (Cuadro 16) manteniendo la homogeneidad de los volúmenes anuales en la media, desviación estándar y el coeficiente de asimetría.

Cuadro 16. Estadísticos de la serie sintética de los volúmenes anuales del río Chancay-Lambayeque (MMC).

Estadísticas	Serie Sintética
Media	1091.54
Desviación estándar	318.92
Coefficiente de variación	0.29
Coefficiente de asimetría	0.11
Coefficiente de autocorrelación	0.11
Sequía más larga	10.00
Máximo déficit	3531.69

Para una muestra dependiente se ha contrastado que la media y desviación estándar de los volúmenes anuales sintéticos se distribuyen asintóticamente a la normalidad $N(0,1)$, (Cuadro 17).

Cuadro 17. Prueba en la media y desviación estándar.

Consistencia		Prueba estadística	
Desviación	F_c	F_t	Significativo
Estándar	0.930	1.130	Si
Media	T_c	T_t	Prueba
	0.361	1.650	Si

De igual modo, la prueba en el coeficiente de correlación evidencia la serie no tuvo una correlación significativa a la tendencia.

Cuadro 18. Prueba en el coeficiente de correlación.

Serie	T_c	T_t	Prueba
Sintética	0.192	1.650	R no es significativo

4.2. ANÁLISIS DE LOS EVENTOS DE SEQUÍA

a. Selección de los niveles de truncamiento

Los umbrales de sequía o nivel de truncamiento de la demanda de agua X_0 se obtuvieron con los coeficientes de déficit de agua λ adoptados igual a 0.85, 0.95, 1.00 y 1.15 del

promedio del volumen anual de la serie observada y la serie sintética generada en la estación Raca Rumi.

Cuadro 19. Nivel de truncamiento de la demanda de agua.

Serie	\bar{X}	λ	X_o
Observada	1091.355	0.85	927.652
		0.95	1036.787
		1.00	1091.355
		1.15	1255.058
Sintética	1091.540	0.85	927.809
		0.95	1036.963
		1.00	1091.54
		1.15	1255.271

De acuerdo con Goldman (1985) los niveles de truncamiento establecidos como la fracción del volumen medio reflejarán la demanda esperada del suministro de agua disponible de la estación Raca Rumi, como el hecho de la ocurrencia de sequías presentes dependen de la interacción de la ocurrencia natural del agua y la intensión de uso operativo. Los volúmenes por debajo por cortos periodos serian sequías intensas corta duración donde la demanda es pequeña y de poca importancia por el amplio almacenamiento del reservorio Tinajones o volúmenes levemente de bajo de largo periodos, son sequías de baja intensidad y larga duración donde usos importantes dependen del almacenamiento (agrícola y poblacional principalmente) pero de poca importancia usuarios de pequeñas fracciones, aun así Fernández (1985); Ashok y Vijay (2010) sustentan que la demanda de agua están sujetas a las características estacionales de la región y a menudo no se encontrarían en fase.

b. Separación de rachas o corridas.

Las rachas o corridas se obtuvieron de la comparación de los volúmenes anuales (observados y sintéticos) con la intención de uso de agua (nivel de truncamiento). La secuencia de anomalías positivas y/o negativas determinaron la serie discreta de la sucesión de años seco (0) y húmedo (1) y los déficits acontecidos d_j ; los eventos de sequía de un déficit hídrico D_o y duración l_o determinada.

Tanto la separación del déficit d_j acontecido como la sucesión de años secos y húmedos resultan una dificultad cuando no se logra percibirlos íntegramente en su tiempo, es decir su ocurrencia dependen del escurrimiento y la intención de uso del agua (Goldman 1985), por ejemplo la representación de sequías que ocurren después de varios meses con caudales insuficientes, recuperándose luego del inicio de la temporada de lluvias en la parte alta de la cuenca, son sequías de una pequeña fracción de agua poco significativo para usos que dependen del almacenamiento, no obstante, el interés puesto en su impacto y como combinarlo con el nivel crítico de uso del agua resulta válido para su estudio.

c. Muestreo del déficit de agua (déficit)

Durante el periodo 1970 a 2019 en la estación Raca Rumi, se ha identificado por nivel de truncamiento hasta ocho eventos de sequía de un año de duración, mientras aquellos de dos, tres y cuatro años solo se observaron una o dos y hasta tres veces.

Cuadro 20. Eventos de sequía ocurridos Do de un año de duración lo.

λ	X_o	lo	Do	Periodo
0.85	927.6517	1	-430.2449	1985
0.85	927.6517	1	-63.2779	1995
0.85	927.6517	1	-265.3399	1997
0.85	927.6517	1	-19.4639	2016
0.95	1036.7872	1	-226.9812	1995
0.95	1036.7872	1	-429.0432	1997
0.95	1036.7872	1	-45.4542	2014
0.95	1036.7872	1	-183.1672	2016
0.95	1036.7872	1	-62.0432	2018
1	1091.3549	1	-40.0049	1970
1	1091.3549	1	-208.7982	1972
1	1091.3549	1	-214.9682	1974
1	1091.3549	1	-31.3182	2000
1	1091.3549	1	-263.7252	2014
1	1091.3549	1	-401.4382	2016
1	1091.3549	1	-280.3142	2018

1.15	1255.0582	1	-203.7082	1970
1.15	1255.0582	1	-45.0949	1972
1.15	1255.0582	1	-51.2649	1974
1.15	1255.0582	1	-281.5489	1995
1.15	1255.0582	1	-483.6109	1997
1.15	1255.0582	1	-100.0219	2014
1.15	1255.0582	1	-237.7349	2016
1.15	1255.0582	1	-116.6109	2018

Los eventos de sequía de un año de duración fueron agrupados según sus cuartiles, para ser contrastados en la estimación de la probabilidad de excedencia de estos.

Cuadro 21. Eventos de sequía observados.

Cuartil %	Valor (MMC)
<25%	19.464
25%	59.3485
50%	206.254
75%	269.0825
>75%	483.611

El déficit acumulado D_0 de duración lo igual a 1, 2, 3 y 4 años de la serie sintética fueron agrupados según su coeficiente de déficit λ de 0.85, 0.95, 1.00 y 1.15. Asimismo, se ha contrastado con la prueba de Grubbs los valores atípicos no han sido significativos. El Cuadro 22 corresponde a la prueba estadística aplicada a los déficits generados de un año de duración.

Cuadro 22. Prueba de Grubs a los déficits generados.

λ	G	G crítico	Significativo
0.85	3.063	3.423	no
0.95	3.054	3.359	no
1.00	2.649	3.132	no
1.15	3.173	3.332	no

4.3. ESTIMACIÓN DE LA PROBABILIDAD DE EXCEDENCIA Y RIESGO DE LOS EVENTOS DE SEQUÍA

a. Transición de sequías

Las series discretas de runs permitió identificar la ocurrencia de un año húmedo siendo el anterior seco, es decir a través del recuento y la frecuencia de un número de transiciones se calculó la probabilidad p_{01} de pasar de un estado de sequía (0) a otro sin sequía (1) (Cuadro 23).

Cuadro 23. Estimación de las probabilidades de transición de los estados.

Serie	λ	X_0	p_{01}	p_{10}	p_0	p_1
Generada	0.85	927.8054	0.6659	0.2992	0.310	0.690
	0.95	1036.9590	0.5052	0.4058	0.445	0.555
	1.00	1091.5358	0.2874	0.6458	0.692	0.308
	1.15	1255.2662	0.4395	0.4637	0.513	0.487

Las probabilidades de transición son las que equilibran el pronóstico probabilístico del estado próximo basado en el estado actual (Avilés 2017), es decir determinan la existencia o no de años de sequía. Se tuvo por encima del 50% una mayor probabilidad p_{01} de pasar de una situación seca a una húmeda resulto para los coeficientes de déficit 0.85 y 0.95 por el contrario 1.00 y 1.15.

b. Ajuste a una distribución de probabilidad bivariado de sequía

La serie hidrológica del déficit de sequía para su ajuste a una PDF Gamma fueron contrastados con la prueba de Anderson-Darling, superando los valores críticos establecidos en los coeficientes de déficit 0.85, 1.00 y 1.15.

Cuadro 24. Prueba de Anderson – Darling a los déficits generados.

λ	AD	p-valor	AD critico
0.85	0.570	0.159	0.784
0.95	0.866	0.032	0.780
1.00	0.420	> 0.25	0.775
1.15	0.596	0.135	0.779

El ajuste de la serie sintética de déficits resultó no ser tan buena para la categoría $\lambda = 0.95$ con un p-valor menor que el nivel de significancia, 0.05. Similar condición presentó (Bellamy *et al.* 2013) para series de déficit reconstruidos en el estudio de análisis de la frecuencia y riesgo de sequías en la parte superior de la cuenca del río Verde, Wyoming-Colorado, debido que la prueba en consideración resulta más potente en el estadístico AD por el mayor peso a las colas de la distribución y el uso de todos los datos en conjunto para el contraste de hipótesis.

c. Estimación de la probabilidad de excedencia de los eventos de sequía acontecidos.

Por el método de máxima verosimilitud se determinó los parámetros de escala γ y forma β de la PDF Gamma bivariado propuesta por Salas *et al.* (2005) (Cuadro 25).

Cuadro 25. Parámetros de la función de distribución de probabilidad Gamma.

Serie	λ	X_0	l_0 (años)	γ	β
	0.85	927.8054	1	1.1094	157.6108
Sintética	0.95	1036.9590	1	1.3383	161.8323
	1.00	1091.5358	1	1.3743	161.8739
	1.15	1255.2662	1	1.6130	173.3152

d. Periodo de retorno y riesgo de sequía

La probabilidad que ocurra un déficit D superior a los déficits D_0 observados de un año de duración, $l_0 = 1$ año, es decir $P[E] = P(D \geq D_0, l_0 = 1 \text{ año})$ categorizados según su coeficiente de déficit en la estación Raca Rumi resultado del producto de la distribución Gamma de dos parámetros del déficit acumulado D_0 y el cálculo de la duración del evento probable de un año húmedo siendo el anterior seco p_{01} (CEDEX 2017; Jackson 1975; Fernández y Salas 1999b).

El cuadro 26 y la figura 13 muestran los cambios en las probabilidades de un déficit D superior o igual a los déficits observados D_0 de un año de duración en la estación Raca Rumi.

Cuadro 26. Probabilidad, periodo de retorno y riesgo de los eventos de sequía.

λ	Do	lo (años)	P[E]	T (años)	R%
0.85	19.4640	1	0.2030	24	4.19
0.85	59.3485	1	0.1634	30	3.37
0.85	206.2540	1	0.0690	70	1.42
0.85	269.0863	1	0.0472	103	0.98
0.85	483.6110	1	0.0124	391	0.26
0.95	19.4640	1	0.2385	19	5.37
0.95	59.3485	1	0.2054	22	4.62
0.95	206.2540	1	0.1016	44	2.29
0.95	269.0863	1	0.0730	61	1.64
0.95	483.6110	1	0.0221	201	0.50
1.00	19.4640	1	0.1963	26	3.90
1.00	59.3485	1	0.1705	29	3.39
1.00	206.2540	1	0.0860	58	1.71
1.00	269.0863	1	0.0622	81	1.24
1.00	483.6110	1	0.0191	264	0.38
1.15	19.4640	1	0.2417	18	5.45
1.15	59.3485	1	0.2217	20	5.00
1.15	206.2540	1	0.1327	33	2.99
1.15	269.0863	1	0.1019	43	2.30
1.15	483.6110	1	0.0378	117	0.85

Es predecible, los periodos de retorno estimados para la estación Raca Rumi fueron de mayor número de años para los eventos de déficit de sequía menos probable en cada una de sus categorías del coeficiente de déficit λ , dado que es más incierto un déficit de sequía presentó uno menor.

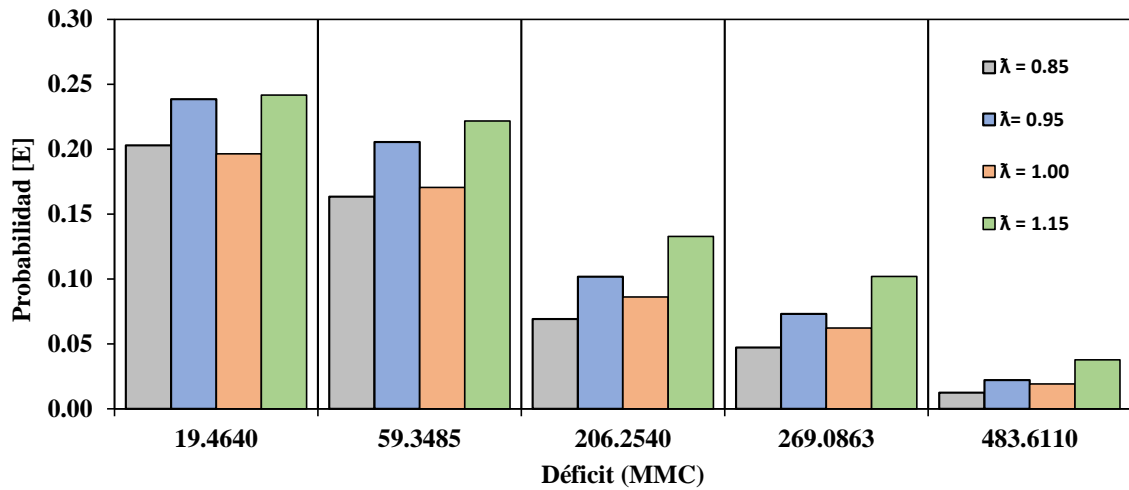


Figura 13. Probabilidad de los eventos de déficit de sequía MMC de la muestra generada del río Chancay Lambayeque respecto a los déficits observados D_o .

Asimismo, se obtuvo los riesgos presenta eventos de déficit de sequía superiores a 483.6110 MMC (0.26% - 0.85%); 269.0863 MMC (0.98% - 2.3%); 206.2540 (1.42% - 2.99%); 59.3485 MMC (3.37% - 5.00%); y 19.4640 MMC (4.19% - 5.45%) entre las categorías de los coeficientes de déficit (Figura 15).

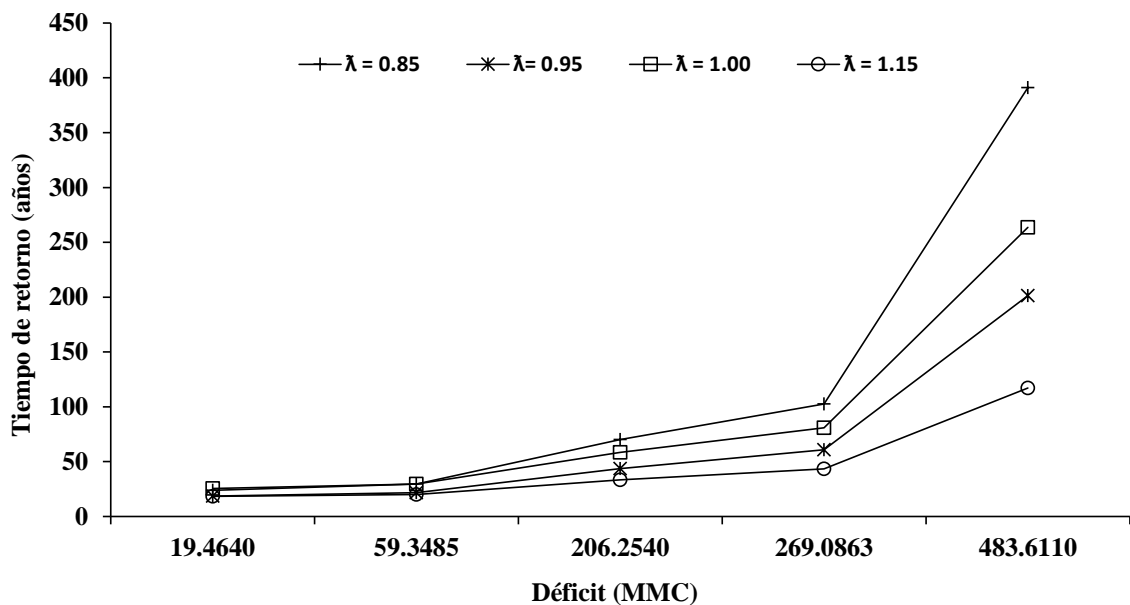


Figura 14. Periodo de retorno de los eventos de déficit de sequía de la muestra generada del río Chancay Lambayeque respecto a los déficits observados D_o .

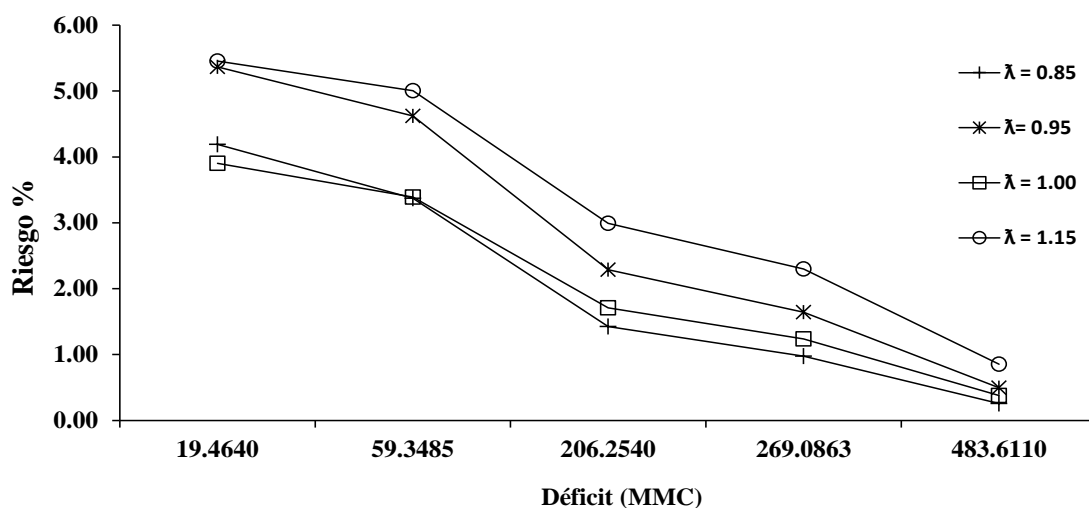


Figura 15. Riesgo de los eventos de déficit de sequía MMC de la muestra generada del río Chancay Lambayeque respecto a los déficits observados Do.

Las pérdidas o daños ocasionados por el déficit de sequía hidrológica en la estación Raca Rumi son pocos conocidos, no se contó con un reporte claro sobre los usos del agua poblacional y agrícola y el tipo de sequía, por ejemplo SEMAHI (s.f) menciona la sequía de 1967-1968 la mas aguda en los últimos 50 años, afecto a los departamentos de Piura, Lambayeque, La Libertad con 125068 hectáreas de cultivos. La ANA (2013) a través del plan participativo de gestión de recursos hídricos de la cuenca Chancay Lambayeque, reportó 109 660 has de riego, 36 centros poblados y 160 255 viviendas afectadas por las sequías.

Es conveniente destacar que la frecuencia y el riesgo de los eventos de déficit de sequías calculados respecto a los déficits observados de un año de duración, son estimaciones con una incertidumbre asociada a la tendencia a largo plazo de la serie de tiempo de los volúmenes de agua natural, las intervenciones humanas con las derivaciones del río Chontano y Conchano al río Chancay Lambayeque, la influencia del fenómeno del Niño y la Niña en sus fase calida y fría, , la dependencia de los volúmenes anuales en el tiempo por la transferencia de agua de un periodo a otro en el sistema (Fernández 1985), el modelo temporal de los volúmenes anuales que mejor los representan (Salas 2000), la periodicidad dominante de la demanda de agua (Ashok y Vijay 2010) y el desfase de la oferta y la demanda de agua (Fernández 1985), hacen que no se mantenga la relación causa-efecto con frecuencia de ocurrencia de sequías, a veces no se pueda detectar y siquiera pueda comprobarse cuando no existen registros válidos de los déficit.

El rango de variación de 29% y el coeficiente de autocorrelación de retardo 1 de 0.068 de los volúmenes anuales de la Estación Raca Rumi muestran que la secuencia de volúmenes anuales es aleatoria sin persistencia. Los cambios bruscos de los volúmenes del río asociados por el trasvase de las aguas como el fenómeno climático del Niño son características importantes de la serie de tiempo (Salas 2000), de manera que tienen implicaciones en la definición de la sequía basada en un umbral.

Asimismo, su componente periódica debido al ciclo anual de los volúmenes, están concentrados en intervalos de tiempo menores, en unos pocos meses del año, lo cual conlleva a tener descargas altas o bajas repentinamente como los eventos extremos de sequía en los periodos secos y las inundaciones en los periodos húmedos. De tal forma las bajas o inexistentes secuencias que exhiben los volúmenes anuales implicaría este río no deberían ocurrir sequías largas y serían otros factores los causantes de sequías extremas en la zona de estudio (Salas 2000).

Por una parte, se conoce que estas sequías hidrológicas asociadas con el fenómeno de El Niño (fase cálida) y la Niña (fase fría) fueron caracterizadas a nivel mensual y trimestral en la Estación Raca Rumi por Curi (2017) y Sosa (2016) mediante el índice de caudal estandarizado IQE y el índice de sequía de caudales SDI correlacionadas con las medias móviles de los tres meses de anomalías de la temperatura superficial del mar TSM en la región Niño 3 y 4. Los caudales se caracteriza por eventos moderadamente secos (14.1%) y extremadamente secos (4.3%). Los meses de noviembre y diciembre son eventos moderadamente húmedos y extremadamente húmedos en la Costa Norte. El riesgo más alto con menor probabilidad de ocurrencia en la fase cálida de El Niño fue 1.2% en temporadas de avenida (diciembre -mayo), 1.3% estiaje (junio-noviembre) y 1.5% el año hidrológico (setiembre-agosto). En la fase fría de La Niña, el riesgo es más bajo con mayor probabilidad de ocurrencia, 0.8% las temporadas de avenida (diciembre -mayo), 0.5% estiaje (junio-noviembre) y 0.2% el año hidrológico (setiembre-agosto).

En el proceso del ciclo hidrológico del río Chancay Lambayeque de acuerdo con CEDEX (2017) cuando las precipitaciones a principios de los meses del año más lluvioso, se retardan y se reparten en diferentes procesos, también se produce un retardo en las secuencias de almacenamiento y transferencias de las agua reguladas, generando la reducción del número de sequías, y ocurren menos veces aunque su duración sea mayor con cambios en sus propiedades como desconectarse entre diferentes procesos en el

momento de precipitaciones con mayor magnitud durante un corto intervalo de tiempo. Por consiguiente, los modelos de rachas se encuentran con esta dificultad en la sucesión de cortos periodos húmedos y secos y de escasez lo que deriva en sequías de corta duración para un año.

De acuerdo con Van Loon y Laaha (2015) las sequías hidrológicas están determinadas por la propagación de la sequía meteorológica, a través del ciclo hidrológico y el terrestre, los impactos de sequías están relacionados con la sequía de la humedad del suelo y la sequía hidrológica, ambos dependen del ecosistema, la población y el agua almacenada en la cuenca.

Considerando el estudio de las sequías meteorológicas por Vega (2018) en la región Vertiente del Pacífico Norte VPN y Andes del Norte NAN donde se emplaza la cuenca del río Chancay Lambayeque, los eventos severos de corto plazo ocurrieron máximo durante siete meses, periodo de noviembre 1989 a mayo 1990 y de largo plazo igualmente los años 1992, 1990 y 1980, afectaron la región de la vertiente del pacifico norte y el norte de los andes, dando lugar a la depresión de los volúmenes de agua (SENAMHI 2019b). Esta relación de las anomalías de las precipitaciones, con la TSM en las regiones 1,2 y 3 y 4 del pacífico relacionadas con el ENSO evidenciaron un menor peligro a las sequías en las regiones NVP y NAN. Los patrones diferenciables espaciales son más claros de la sequía de índice de precipitación estandarizada IPE-6 meses en la región.

Estas clasificación pese que ha permitido caracterizar las sequías hidrológicas como las meteorológicas con influencia del Niño o La Niña y distinguir la relación de los IPE o IQE con las TSM, no se identifica respecto a la propagación del déficit de lluvia-recurso y fallo en el sistema de regulación (CEDEX 2017) cuando los eventos de sequía podrían ser superados por un déficit mayor, es decir el régimen del río Chancay Lambayeque se produjo sequía hidrológica cuando el volumen de agua no fue suficiente para proveer una demanda de agua.

Beard y Kubik (1972) examinaron el efecto de distintas sequías con el uso y los requerimientos de suministro del reservorio de los ríos Arroyo Seco de California y Willamette Albania, Orlando, encontraron la mayoría de las sequías de corta duración no son severas cuando no se proporciona poco o ninguna regulación para su almacenamiento, por el contrario, son severas de corta duración durante la temporada de irrigación. Sin

embargo, cuando existe un alto grado de almacenamiento comprometido, las sequías de corta duración son insignificantes. En efecto, en los casos presentados, la regulación del río Chancay Lambayeque como la recuperación del volumen de agua después de un periodo sin lluvias a principios del mes del año de mayor precipitación, los eventos de escasos meses con déficit de sequía no son severos de acuerdo con el riesgo estimado en cada categoría de la demanda de agua. Según Cano (2013) este déficit de sequía en consideración estaría relacionado no por la falta de agua sino por la falta de gestión durante su aprovechamiento con pérdidas hasta del 62%, de acuerdo con el balance hídrico del periodo 2012-2013, la cuenca no tiene déficit hídrico pues recibe de los trasvases un quinto del aporte del caudal del agua del Chancay Lambayeque.

Se ha logrado caracterizar los eventos de déficit de sequía de un año de duración para conocer su severidad y riesgo existentes a diferentes coeficientes de demanda de agua. El análisis será de mayor ayuda si se vence el problema de escasez de datos de la demanda de agua mensual necesarios en regiones críticas como estas afectadas por el incremento de las sequías, el cambio climático y el crecimiento de la demanda de agua (Ashok y Vijay 2010).

Particularmente durante el uso del modelo de rachas la falta de información de la demanda con los volúmenes anuales (oferta) disponibles, dificultó la identificación de la sucesión de eventos húmedos y secos de corto periodo, con sequías cortas, fragmentadas y dependientes, influenciadas por el fenómeno de El Niño y La Niña. Los eventos de déficit de sequía bajo los umbrales de entrada y la recuperación de la sequía en la estación Raca Rumi durante los últimos cincuenta años, revelaron hasta ocho sucesos de un año de duración y escasos o menos de tres sucesos de sequía de dos, tres y cuatro años de duración. Estos sucesos en las series sintéticas a largo plazo pudieron representarse con un mayor número de ocurrencias del déficit de sequía de 1, 2, 3 y 4 años de duración y las ocurrencias de un estado húmedo a uno seco. Sin embargo, la probabilidad de excedencia de los eventos de déficit de sequía de 2, 3 y 4 años de duración no pudo ser estimada debido que no convergen con la sucesión propia de las sequías.

El modelo probabilístico de medias móviles MA (1) para el pronóstico de volúmenes anuales y los déficit de sequía a largo plazo cumplieron con las pruebas requeridas, mientras el modelo de análisis de sequía bivariado propuesto por Salas *et al.* (2005) presento limitaciones en el cálculo de las probabilidades de un año húmedo siendo el

anterior seco, para las secuencias de estados de las sequías de 2,3 y 4 años de duración debiendo estar balanceadas 50%, por el contrario los déficit de sequía de 1, 2, 3 y 4 años de duración si se ajustaron a la distribución Gamma de dos parámetros, lo que limitó la estimación de los eventos de déficit de sequía superiores a 1 año de duración.

Para preservar las estadísticas estacionales como las características de las sequías de los volúmenes del río Chancay Lambayeque en la Estación Raca Rumi y simular los procesos hidrológicos, Salas y Boes (1980) sugieren el modelo “Shifting Level Process” (SLP). Otra alternativa utilizada por CEDEX (2017) en el estudio del impacto de las sequías hidrológicas en España, consistió en simular series sintéticas de volúmenes mensuales de 6000000 meses que conserven sus propiedades estadísticas aplicando un modelo ARMA, así tener una serie más larga y que extiendan la autocorrelación.

Sin embargo, la distribución de probabilidad bivariado del déficit de sequía (Gamma) y la distribución de la secuencia de estados de déficit y superávit (cadena simple de Markov), para el cálculo del periodo de retorno de la sequía hidrológica, presenta el inconveniente de contar la misma familia para cada distribución, y no resulta adecuado cuando la serie exhibe una autocorrelación inferior. Para superar este problema Ashok y Vijay (2010) recomiendan utilizar el modelo de copulas conjunta bidimensionales, tanto en la duración (distribución exponencial) y la severidad (distribución gamma).

Finalmente, para mejorar la fiabilidad del pronóstico a gran escala, se puede incorporar de igual manera con los índices de caudales IQE y los IPE por Sosa (2016) y Vega (2018), determinar los predictores del promedio de las temperaturas superficiales trimestrales del mar SST en la región de El Niño 1 y 2 y la región 3 y 4 con los déficits de sequía simulados y demostrar que un determinando evento de severidad y duración de sequía está ocurriendo. Incluso según Rojas y Anculle (2014) se predeciría el comienzo de la sequía con un mes de anticipación, a través de la relación de las sequías con el estrés hídrico de los cultivos, el índice de vegetación de diferencia normalizada NDVI afectada por los eventos del El Niño 1 y 2 y las temperaturas superficiales trimestrales del mar SST.

V. CONCLUSIONES

En el presente estudio se analizó la severidad y riesgo de sequía del río Chancay Lambayeque mediante la utilidad de los modelos probabilísticos para definir a largo plazo los déficits que se presentan en la estación Raca Rumi y estimar la probabilidad de excedencia y los riesgos que se producen durante el periodo 1970-2019, alcanzando las siguientes conclusiones:

- Se examinó la serie de volúmenes anuales del río Chancay Lambayeque, 1970-2019, de la estación Raca Rumi, identificando se distribuyen normalmente y, confirmando estadísticamente la homogeneidad de la media y la desviación estándar y la estacionariedad de los datos.
- Se determinó el modelo estocástico de medias móviles de orden 1 MA (1) adecuado para representar la estructura de dependencia de la serie de volúmenes anuales del río Chancay Lambayeque, periodo 1970-2019, capaz de reproducir las principales estadísticas y los déficits de sequías simulados a largo plazo.
- Se estimó las probabilidades de excedencia y los periodos de retorno de los déficits de sequía de un año de duración acontecidos para los volúmenes anuales del río Chancay Lambayeque mediante el modelo de distribución de probabilidad bivariado Gamma de los déficits generados y la probabilidad de años secos y húmedos a largo plazo.

VI. RECOMENDACIONES

De acuerdo con los resultados alcanzados en la presente investigación, se tiene las siguientes recomendaciones:

- Usar registros periódicos de la demanda de agua igual a la intención de uso del recurso hídrico del río Chancay Lambayeque.
- Clasificar las sequías mediante técnicas de cuantiles de media móvil mensual, cuantiles de ventanas móviles de treinta días, aproximación rápida de transformación de Fourier entre otros, capaces de obtener las secuencias de escasez de agua.
- Generar series sintéticas aplicando el modelo Shifting Level Process o ARIMA utilizando las series volúmenes mensuales del río Chancay Lambayeque.
- Utilizar el modelo de copulas conjunta bidimensionales para la distribución de la duración y la severidad de las sequías mensuales.
- Realizar investigaciones para estudiar el impacto de cambio climático de las sequías hidrológicas en la cuenca Chancay Lambayeque, incorporando predictores a gran escala de la TSM en las regiones de El Niño 1 y 2 – 3 y 4 con los NDVI y los déficits de sequía.

VII. REFERENCIAS BIBLIOGRÁFICAS

Aliaga Araujo, SV. 1980. Análisis de las características de sequías por el método experimental-teoría del run aplicado a la cuenca del río Chicama. Tesis Magister Scientiae. Lima, Perú, UNALM.

Aliaga, S. 1985. Hidrología Estadística. 1ra Edición. Lima, Perú. 328 p.

Alvarado Medellín, P. 2007. Modelación estocástica de los escurrimientos de la cuenca del Río Amajac. Tesis Dr. Montecillo - Texcoco, México. Instituto de Enseñanza e Investigación en Ciencias Agrícolas. 158 p. <https://1library.co/document/9yner1jy-modelacion-estocastica-escurrimientos-cuenca-rio-amajac-hidalgo-mexico.html>

ANA (Autoridad Nacional del Agua). 2013. Plan participativo de gestión de recursos hídricos en el ámbito del Consejo de Recursos Hídricos de la cuenca Chancay – Lambayeque. Lambayeque. Perú. 183 p. <https://hdl.handle.net/20.500.12543/20>

Ashok, M; Vijay, S. 2010. A review of drought concepts. Journal of Hydrology 391(1–2):202-216. ISSN 0022-1694, <https://doi.org/10.1016/j.jhydrol.2010.07.012>.

Avilés Añazco, AM. 2017. Pronóstico probabilístico de eventos de sequías y evaluación del riesgo en la gestión de sistemas de recursos hídricos, Caso de estudio en una cuenca andina regulada. Tesis Dr. no publicada. Valencia, España. Universidad Politécnica de Valencia. <https://doi.org/10.4995/Thesis/10251/78456>

Bastidas Valdivia, JM. 2015. Generación de descargas en la cuenca del río Mantaro mediante el uso de modelos estocásticos de series sintéticas. Tesis pregrado. Huancayo, Perú. UNCP. 261 p.

Beard, LR; Kubik, HE. 1972. Droughts Severity and Water Supply Dependability. Technical Paper N° 30, Hydrologic Engineering Center, Corps of Engineers, U.S. Army, Davis, California. 19 p.

Bellamy, JT; Tootle, G; Huzurbazar, S; Pochop, L; Barnett, A. 2013. Case Study of Drought Frequency and Risk Analysis in the Upper Green River Basin, Wyoming. Journal of Hydrologic Engineering 18:888-896.

Cano, A. 2013. ¿«De arriba hacia abajo» o «de abajo hacia arriba»? : participación social, agricultura y minería en la gestión integrada de la cuenca Chancay-Lambayeque. Apuntes. Revista De Ciencias Sociales 40(73):43-76. <https://doi.org/https://doi.org/10.21678/apuntes.73.687>

Castillo, S; Lozano, ED. 2007. Q-Q Plot Normal. Los puntos de posición gráfica. Iniciación a La Investigación. 2:1-20. <https://revistaselectronicas.ujaen.es/index.php/ininv/article/view/259>

CEDEX. 2017. Evaluación del impacto del cambio climático en los recursos hídricos y sequías en España. Informe Técnico para Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente. Madrid, España. 320 p.

Curi Tapahuasco, S. 2017. Caracterización de de las sequías hidrológicas en la vertiente peruana del Océano Pacífico. Tesis Magister Scientiae. Lima, Perú, UNALM. 126 p.

Chow, VT; Maidment, DR; Mays, LW. 1988. Hidrológica Aplicada. Santa Fe de Bogotá. Colombia, Mac Graw Hill. 84 p.

CLD. 2004. Convección de las Naciones Unidas de lucha contra la desertificación en los países afectados por la sequía grave o la desertificación en particular en África. 57 p. [WWW Document]. URL. https://www.unccd.int/sites/default/files/relevant-links/2017-08/UNCCD_Convention_text_SPA.pdf (accessed 20.08.2021).

Dracup, JA; Seong, LK; Paulson, EG. 1980. On the statistical characteristics of drought events. Water Resources Research. 16(2):289-296.

Fernández, B; Salas, JD. 1999a. Return Period and Risk of Hydrologic Events. I: Mathematical Formulation. Journal of Hydrologic Engineering 4(4):297-307.

Fernández, B.; Salas, JD. 1999b. Return Period and Risk of Hydrologic Events. II: Applications Journal of Hydrologic Engineering 4(4):308-316.

Fernández, B. 1985. Análisis de la periodicidad en series hidrológicas. Anales de la Universidad de Chile 8:635-653.

Fernández, B. 1997. Identificación y caracterización de sequías hidrológicas en Chile central. Ingeniería del Agua 4(4):37-46.

- Goldman, D. 1985. Stochastic Analysis of Drought Phenomena. US Army Corps of Engineers. Institute for Water Resource. HEC. TD-25, 140 p.
- Heo, JH; Kho, YW; Shin, H; Kim, S; Kim, T. 2008. Regression Equations of Probability Plot Correlation Coefficient Test Statistics from Several Probability Distributions. Journal of Hydrology 335(1-4):1-15.
- Jackson, BB. 1975. "Markov mixture models for drought lengths". Water Resour, 11(1):64–74.
- Linsley, RK; A Kholer, M; Paulhus, J. (1949). Applied hydrology, SERBIULA (Sistema Librum 2.0). McGraw-Hill Book Co.
- Maidment, D. 1993. Handbook of Hydrology. Cap 19 Analysis Series Times. Mc Graw Hill Professional. 1424 p.
- OMM. (Organización Mundial de la Salud). 1992. La Conferencia Mundial sobre el Clima. Convención marco de las Naciones Unidas sobre el cambio climático. Nueva tirada por razones técnicas. 26 p.
<https://www.minambiente.gov.co/index.php/component/content/article/420-plantilla-cambio-climatico-10>
- Pankratz., A. 1983. Forecasting with univariate Box-Jenkins models, concepts, and cases, John Wiley and Sons Inc. New York, USA.
- Peña, SD. 2010. Análisis de Series temporales. España. Alianza Editorial S. A. 604 p.
- PEOT. (Proyecto Especial Olmos Tinajones). 2004. Sistema Tinajones [Power Point]:
<https://www.regionlambayeque.gob.pe/web/tema/detalle/4435?pass=MTA1Nw==>
- PEOT. (Proyecto Especial Olmos Tinajones). 2011. Plan operativo anual 2011. 25 p.
https://www.peru.gob.pe/docs/PLANES/13473/PLAN_13473_Plan_Operativo_Institucional_-_POI_2011.pdf
- Pérez, RF. 2010. Modelo ARIMA-ARCH Algunas aplicaciones de las series de tiempo financieras. Colombia. Sello Editorial Universidad de Medellín. 96 p.
- Pérez, RF. 2012. Métodos no paramétricos para el análisis de series de tiempo. Colombia, Sello Editorial. Universidad de Medellín. 89 p.

Rojas, J; Anculle, T. (19 – 21 de noviembre 2014). Predicción del comienzo de la sequía en la costa norte de Perú usando el ndvi, datos in-situ y los datos el niño oscilación sur (1982-2013) [Discurso Principal]. Simposio Internacional de Expertos Manejo de Sequías, Santiago, Chile.

Sakia, RM. 1992. The Box-Cox transformation technique: A review, *Statistician* 41:169-178. doi:10.2307/2348250.

Salas, JD. 2000. Hidrología de zonas áridas y semiáridas, *Ingeniería del Agua* 7(4):409-428.

Salas, JD; Boes, DC. 1980. Shifting level modelling of hydrologic series. *Advances in Water Resources*, 3(2):59-63. doi:[10.1016/0309-1708\(80\)90028-7](https://doi.org/10.1016/0309-1708(80)90028-7)

Salas, JD; Delleur, JW; Yevjevich, V; Lane, WL. 1985. Applied Modeling of hydrologic Times Series. *Water Resources Research* 33(2):291-308.

Salas, JD; Fu, CJ; Cancelliere, A; Dustin, D; Bode, D; Pineda, A; Vincent, E. 2005. Characterizing the severity and risk of drought in the Poudre River, Colorado. *Journal of Water Resources Planning and Management-ASCE* 131(5):383-393.

Salvador, F; Gargallo, P. 2003. Análisis exploratorio de datos. [en línea] 5 campus.com, Estadística <http://www.5campus.com/leccion/aed> [Julio, 2007].

Schneider, SH; Mastrandrea, M; Root, TL. 2011. *Encyclopedia of Climate and Weather*, OUP USA.

SENAMHI. 2019a. Monitoreo de sequías hidrológicas: Condiciones de déficit hídrico a enero 2019 en la Vertiente del Pacífico (norte). Lima, Perú.

SENAMHI. 2019b. Caracterización Espacio Temporal de la Sequía en los Departamentos Altoandinos del Perú (1981-2018). Lima, Perú.

SENAMHI, s.f. Aspectos de las sequías en el Perú. Lima, Perú.

Shiau, JT; Shen, HW. 2001. Recurrence analysis of hydrologic droughts of differing severity. *Journal of Water Resources Planning and Management-ASCE* 127(1):30-40.

Sosa Sarmiento, JD. 2016. Análisis de la sequía hidrológica en el Perú. Tesis pregrado. Lima, Perú. UNALM. 197 p.

Sveinsson, O; Salas, J; Lane, W; Frevet, D. 2007. Stochastic Analysis, Modelling, and Simulation (SAMS) User Manual. Technical Report N° 11. University State Colorado. 118 p.

Van Loon, AF; Laaha, G. 2015. Hydrological drought severity explained by climate and catchment characteristics. Journal of Hydrology, 526:3-14.

Vega Jácome, F. 2018. Variabilidad espacio temporal de las sequías en el Perú y el peligro asociado al fenómeno del niño. Tesis Magister Scientiae. Lima, Perú, UNALM. 96 p.

Vega, Y. (Diciembre, 3, 2020). Lambayeque: Reservorio Tinajones con 93 millones de metros cúbicos de agua. <https://larepublica.pe/sociedad/2020/12/03/lambayeque-reservorio-tinajones-con-93-millones-de-metros-cubicos-de-agua-lrnd/>

Yevjevich, VM. 1972. Stochastic Processes in hydrology, Water Resources Publication, Colorado State University, Fort Collins.

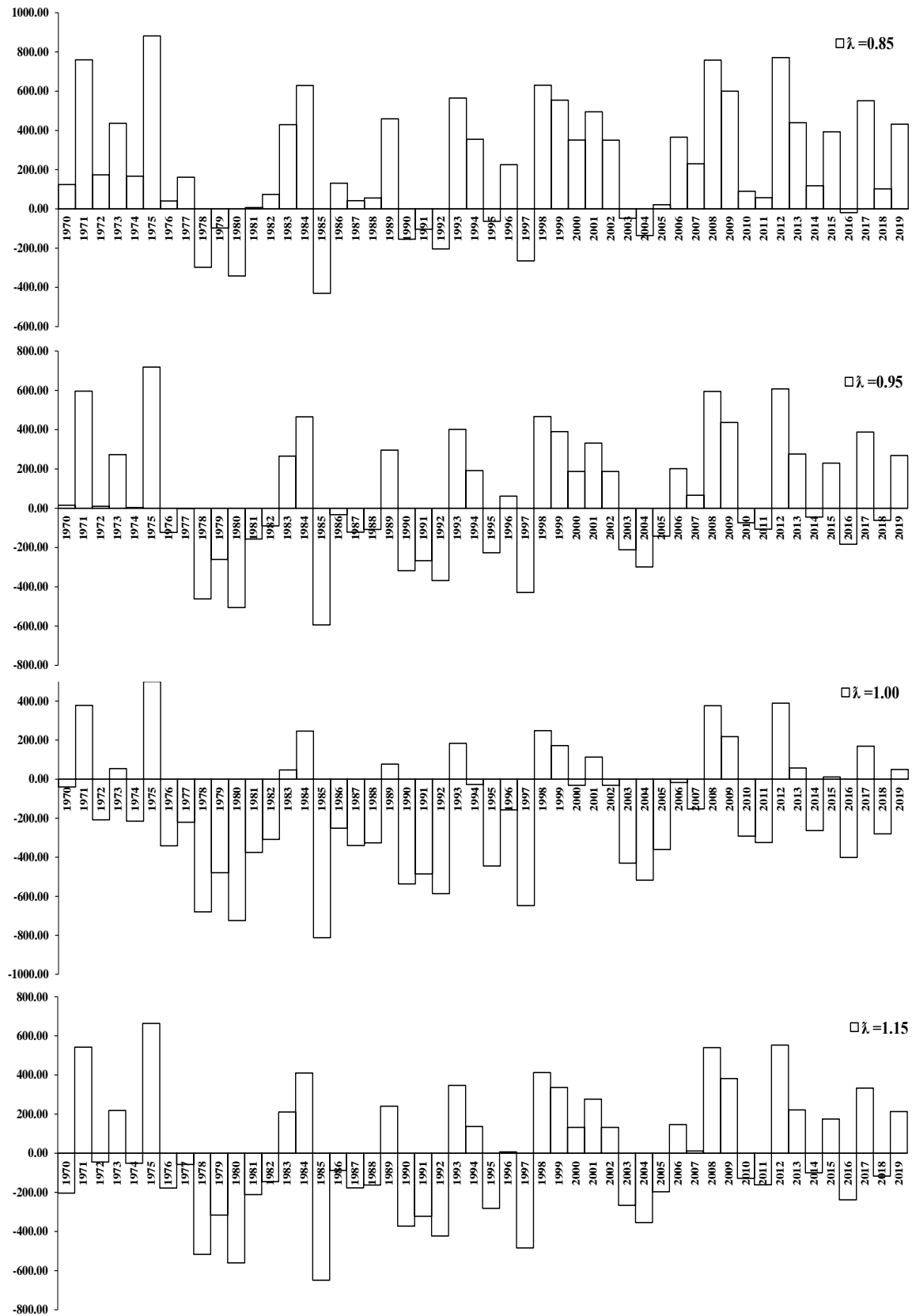
VIII. ANEXOS

Anexo 1. Serie histórica de volúmenes anuales de descargas del río Chancay Lambayeque. Estación Raca Rumi en millones de metros cúbicos 1914 – 2019.

Año	X _t	Año	X _t	Año	X _t
1914	999.6480	1950	497.9540	1986	1003.7784
1915	1069.3728	1951	544.3032	1987	914.6574
1916	774.6451	1952	664.6584	1988	928.4643
1917	1374.5376	1953	1042.3940	1989	1332.0183
1918	1051.3843	1954	670.8956	1990	718.6850
1919	1096.5024	1955	657.4961	1991	769.6520
1920	929.9359	1956	831.0060	1992	668.7222
1921	1146.9626	1957	747.9525	1993	1438.1453
1922	1229.9472	1958	685.1473	1994	1227.9885
1923	855.3254	1959	698.3108	1995	809.8056
1924	825.9408	1960	710.1716	1996	1098.1189
1925	3260.2522	1961	644.1761	1997	607.7440
1926	783.6853	1962	823.5581	1998	1503.2657
1927	968.7985	1963	476.9645	1999	1426.9340
1928	1070.6372	1964	913.7632	2000	1223.7390
1929	1125.7307	1965	820.1667	2001	1368.0591
1930	1105.2634	1966	663.5311	2002	1223.4137
1931	717.7686	1967	808.1237	2003	825.0470
1932	1042.6647	1968	468.8556	2004	737.5783
1933	1000.5593	1969	951.1582	2005	894.4270
1934	850.2329	1970	1051.3472	2006	1238.2390
1935	726.0587	1971	1632.9164	2007	1102.8700
1936	788.5333	1972	1046.2567	2008	1631.3555
1937	493.9222	1973	1309.4850	2009	1473.1993
1938	792.1757	1974	1040.0927	2010	963.0731
1939	821.0999	1975	1754.8861	2011	930.0384
1940	698.9820	1976	913.4071	2012	1643.9936
1941	1046.3570	1977	1034.5216	2013	1312.4786
1942	664.6584	1978	575.1003	2014	991.3332
1943	826.1044	1979	775.3781	2015	1266.1789
1944	763.8059	1980	531.1594	2016	853.6197
1945	685.7959	1981	879.8069	2017	1424.1918
1946	813.0456	1982	946.5959	2018	974.7435
1947	968.4568	1983	1302.0953	2019	1304.7234
1948	822.4150	1984	1501.5610		
1949	650.0358	1985	442.8395		

Fuente: Proyecto Especial Olmos Tinajones (2004).

Anexo 2. Serie histórica de rachas de sequía por coeficientes de déficit $g=0.85, 0.95, 1.00$ y 1.15



Anexo 3. Estadísticos de 10 series generadas de 50 años cada una en millones de metros cúbicos MA (1)

Estadísticos	Xgen1	Xgen2	Xgen3	Xgen4	Xgen5	Xgen6	Xgen7	Xgen8	Xgen9	Xgen10
Media	1061.550	1069.700	1089.310	1100.500	1083.280	1101.390	995.8430	1120.750	1014.170	1046.990
Desv. Estand.	0	0	0	0	0	0	0	0	0	0
Asimetría	299.1420	267.9170	299.9050	349.0510	325.4600	318.4510	256.1370	289.0230	347.9670	289.9050
Mínimo	0.2592	0.3897	0.1866	0.3349	0.4450	0.1162	0.0502	0.0185	0.1079	-0.7930
Máximo	335.2610	568.7200	483.9320	396.6140	464.7170	404.1980	385.5000	486.6030	245.1940	14.0761
FAC (1)	1835.430	1749.440	1691.820	1973.740	1915.890	1928.620	1577.280	1760.090	2030.230	1764.000
FAC (2)	0	0	0	0	0	0	0	0	0	0
FAC (3)	0.0953	-0.0151	-0.0080	0.2142	0.2078	-0.1175	0.2926	0.3918	-0.0951	0.0364
FAC (4)	-0.0994	0.0788	-0.0016	0.1543	0.0593	-0.0814	-0.0425	-0.0104	-0.1827	-0.0767
FAC (5)	0.0371	0.0826	0.1257	-0.0762	-0.1934	-0.0728	-0.0867	-0.2094	0.1180	-0.1136
Nivel Demanda	0.1172	0.1332	-0.0913	0.0851	0.0613	-0.0997	0.0740	-0.1707	-0.2539	-0.1667
Sequía más larga	0.1245	0.0286	0.0527	0.0910	-0.2450	0.1976	-0.0088	-0.0831	0.0672	0.1052
Máximo déficit	1091.350	1091.350	1091.350	1091.350	1091.350	1091.350	1091.350	1091.350	1091.350	1091.350
Superávit más largo	0	0	0	0	0	0	0	0	0	0
Máximo superávit	8.0000	6.0000	6.0000	7.0000	4.0000	4.0000	12.0000	4.0000	4.0000	5.0000
Capacidad de almacenamiento	1794.490		1763.800	1241.270	1614.650	1133.480	3003.860	1434.820	1654.770	1752.580
Rango rescalado	0	998.4880	0	0	0	0	0	0	0	0
Coeficiente de Hurst	6.0000	3.0000	5.0000	6.0000	6.0000	5.0000	4.0000	6.0000	2.0000	4.0000
	2060.060	1667.040		2155.470	1952.860	1096.760	1335.300	2481.910		
	0	0	908.4370	0	0	0	0	0	947.8910	906.0840
	3552.040	3202.210	3564.060	2133.330	1614.650	1789.390	5315.860	1434.820	4495.300	2957.340
	0	0	0	0	0	0	0	0	0	0
	8.0098	9.3387	11.7477	10.4102	8.2200	6.5194	10.1030	7.9771	4.6521	7.1407
	0.6464	0.6941	0.7654	0.7278	0.6544	0.5824	0.7185	0.6451	0.4776	0.6107

Anexo 4. Estadísticos de 10 series generadas de 1500 años cada una en millones de metros cúbicos MA (1)

Estadísticos	Xgen1	Xgen2	Xgen3	Xgen4	Xgen5	Xgen6	Xgen7	Xgen8	Xgen9	Xgen10
Media	1094.44	1103.46	1080.68	1090.87	1096.71	1075.37	1091.54	1096.69	1098.64	1095.1
Desv. Estand.	315.401	314.967	318.068	308.371	320.383	308.385	318.918	317.951	319.936	311.525
Asimetría	0.285067	0.162341	0.203595	0.151472	0.214574	0.248242	0.111755	0.147337	0.230075	0.213275
Mínimo	14.0761	31.6497	-37.7699	182.352	65.1216	247.29	202.663	106.416	232.393	77.4085
Máximo	2496.87	2305.91	2323.49	2120.43	2428.74	2066.33	2170.04	2211.24	2137.92	2120.05
FAC (1)	0.105726	0.0880397	0.0613531	0.059622	0.0250715	0.109915	0.113786	0.08662	0.0664295	0.0407854
FAC (2)	-0.020603	0.0231937	-0.0463233	-	0.00294779	-0.0400957	0.0126056	0.00803516	-0.0115919	-
FAC (3)	-0.048403	0.039967	-0.0528914	-0.0106505	-0.0378362	0.0293429	-0.0260108	-0.0557639	0.0193381	0.00722727
FAC (4)	-0.024826	-	0.00554564	0.00428838	0.00944851	-	-	0.00933529	0.0136012	0.065826
FAC (5)	0.010081	-0.0221638	0.039291	0.0471632	-0.0273333	-0.0384946	0.037692	0.0411975	-	-0.0199428
Nivel Demanda	1091.35	1091.35	1091.35	1091.35	1091.35	1091.35	1091.35	1091.35	1091.35	1091.35
Sequía más larga	12	12	11	10	8	11	10	16	11	9
Máximo déficit	3003.86	4082.78	3791.27	2730.17	2301.59	3368.74	3531.69	4636.15	2322.7	2755.96
Superávit más largo	7	11	13	9	10	7	8	12	9	13
Máximo superávit	2481.91	3072.74	4087.01	2692.21	2808.35	3046.83	3261.08	2884.85	2805	4255.25
Capacidad de almacenamiento	10380.8	12851.2	23127.3	16861.7	11249.3	30938.4	14990.7	14002.6	10242.1	10562.1
Rango rescalado	36.4221	72.3875	61.4675	53.9996	49.0941	58.9161	47.6409	54.0164	59.2167	37.941
Coefficiente de Hurst	0.543072	0.646826	0.622124	0.602558	0.588172	0.615721	0.583633	0.602605	0.616489	0.549243

Anexo 5. Prueba estadística de Grubbs

Es una prueba estadística para detectar valores atípicos de un conjunto de datos univariados que provienen de una población distribuida normalmente.

La técnica consiste en detectar un valor atípico por cada realización, este valor se elimina del conjunto de datos y el contraste se repite hasta no detectar valores atípicos.

La prueba de Grubbs tiene como hipótesis:

Ho: no hay valores atípicos en el conjunto de datos

Ha: hay un valor atípico (mínimo o máximo) en el conjunto de datos

El estadístico de la prueba de Grubbs es:

$$G = \frac{|X_{atípico} - \bar{X}|}{S}$$

Donde: \bar{X} y S son la media y desviación estándar de los valores, $X_{atípico}$ valor anómalo sospechoso.

El estadístico de la prueba de Grubbs es la mayor desviación absoluta de la media muestral en unidades de la desviación estándar muestral. El estadístico G se compara con un valor crítico de G según el tamaño de la muestra y el nivel de significancia seleccionado. Si G es mayor a G crítico, la hipótesis nula es rechazada y el dato contrastado es un outlier.

El Cuadro 27 muestra los valores críticos para un 95% y 99% de nivel de confianza y tamaño de la muestra N.

Cuadro 27. Valores críticos de la prueba de Grubbs.

N	G _{crítico} 95%	G _{crítico} 99%	N	G _{crítico} 95%	G _{crítico} 99%
3	1.15	1.15	19	2.53	2.85
4	1.46	1.49	20	2.56	2.91
5	1.67	1.75	21	2.58	2.94
6	1.82	1.94	22	2.60	2.96
7	1.94	2.10	23	2.62	2.99
8	2.03	2.22	24	2.64	3.01
9	2.11	2.32	25	2.66	
10	2.18	2.41	30	2.75	
11	2.23	2.48	35	2.82	
12	2.29	2.55	40	2.87	
13	2.33	2.61	45	2.92	
14	2.37	2.66	50	2.96	
15	2.41	2.71	60	3.03	
16	2.44	2.75	70	3.09	
17	2.47	2.79	80	3.14	
18	2.50	2.82	90	3.18	
19	2.53	2.85	100	3.21	

Anexo 6. Prueba estadística de Anderson – Darling para una PDF Gamma

La prueba de bondad de ajuste de Anderson -Darling medida sobre los datos de una muestra ordenados de forma creciente $x_1 \leq x_2 \leq \dots \leq x_n$ contrasta si se ajustan a una distribución continua con función de distribución acumulativa $F(x)$.

Las hipótesis para la prueba AD son:

Ho: los datos provienen de una distribución especificada.

Ha: Los datos no provienen de una distribución especificada.

Donde: n el tamaño de la muestra, $F(x)$ = CDF para la distribución especificada, i = la i muestra, calculada cuando los datos se ordenan en orden ascendente.

$$A = -n - \frac{1}{n} \sum_{i=1}^n [(2i - 1)[LNF(x_i) + LN(1 - F(X_{n-i+1}))]]$$

Cuando no se conocen los parámetros de la distribución, pero deben estimarse a partir de la muestra, debemos realizar algunos pasos adicionales. Primero, calculamos el estadístico de prueba AD basado en A , pero que depende de la distribución específica, como se describe en el cuadro 28.

En el Cuadro 28 Estadístico AD según la distribución adoptada

Distribución	AD valor
Gamma ($k=1$)	$A \left(1 + \frac{0.60}{n}\right)$
Gamma ($k \geq 2$)	$A + \frac{0.20 + 0.30/k}{n}$

Determinamos que existe una diferencia significativa entre los datos de la muestra y la distribución teórica proporcionada $AD \geq AD_{crit}$, donde AD_{crit} es el valor crítico (para un valor dado de α) definido en base a la tabla de valores críticos.

Cuadro 28. Valores críticos para la prueba Anderson-Darling a una distribución Gamma.

k/a	0.005	0.010	0.025	0.050	0.10	0.25
1	1.227	1.092	0.917	0.786	0.657	0.486
2	1.190	1.062	0.894	0.768	0.643	0.477
3	1.178	1.052	0.886	0.762	0.639	0.475
4	1.173	1.048	0.883	0.759	0.637	0.473
5	1.170	1.045	0.881	0.758	0.635	0.472
6	1.168	1.043	0.880	0.757	0.635	0.472
8	1.165	1.041	0.878	0.755	0.634	0.471
10	1.164	1.040	0.877	0.754	0.633	0.471
12	1.162	1.038	0.876	0.754	0.633	0.471
15	1.162	1.038	0.876	0.754	0.632	0.470
20	1.161	1.037	0.875	0.753	0.632	0.470
>	1.159	1.035	0.873	0.752	0.631	0.470

Anexo 7. Déficit observado en la estación Raca Rumi 1970-2019.

λ	Xo	lo	Do	Periodo(s)
0.85	927.6517	1	-430.2449	1985
0.85	927.6517	1	-63.2779	1995
0.85	927.6517	1	-265.3399	1997
0.85	927.6517	1	-19.4639	2016
0.85	927.6517	2	-183.5429	2003-2004
0.85	927.6517	3	-737.6148	1978-1979-1980
0.85	927.6517	3	-462.1928	1990-1991-1992
0.95	1036.7872	1	-226.9812	1995
0.95	1036.7872	1	-429.0432	1997
0.95	1036.7872	1	-45.4542	2014
0.95	1036.7872	1	-183.1672	2016
0.95	1036.7872	1	-62.0432	2018
0.95	1036.7872	2	-180.4633	2010-2011
0.95	1036.7872	3	-953.3025	1990-1991-1992
0.95	1036.7872	3	-653.3095	2003-2004-2005
0.95	1036.7872	4	-857.4087	1985-1986-1987-1988

1	1091.3549	1	-40.0049	1970
1	1091.3549	1	-208.7982	1972
1	1091.3549	1	-214.9682	1974
1	1091.3549	1	-31.3182	2000
1	1091.3549	1	-263.7252	2014
1	1091.3549	1	-401.4382	2016
1	1091.3549	1	-280.3142	2018
1	1091.3549	2	-617.0053	2010-2011
1	1091.3549	3	-1608.1155	1990-1991-1992
1	1091.3549	4	-1730.4926	1985-1986-1987-1988
1	1091.3549	4	-1276.5726	1994-1995-1996-1997
1.15	1255.0582	1	-203.7082	1970
1.15	1255.0582	1	-45.0949	1972
1.15	1255.0582	1	-51.2649	1974
1.15	1255.0582	1	-281.5489	1995
1.15	1255.0582	1	-483.6109	1997
1.15	1255.0582	1	-100.0219	2014
1.15	1255.0582	1	-237.7349	2016
1.15	1255.0582	1	-116.6109	2018
1.15	1255.0582	2	-289.5988	2010-2011
1.15	1255.0582	3	-1117.0058	1990-1991-1992
1.15	1255.0582	3	-817.0128	2003-2004-2005
1.15	1255.0582	4	-1075.6797	1985-1986-1987-1988

Anexo 8. Resumen de la prueba de ajuste Anderson Darling y estimación de parámetros PDF Gamma para los déficits generados.

lo	λ	AD	p-valor	AD crítico	γ	β	Prueba
1	0.85	0.5696	0.1587	0.7840	1.1094	157.6108	Ajusta PDF Gamma
1	0.95	0.8658	0.0315	0.7799	1.3383	161.8323	Ajusta PDF Gamma
1	1.00	0.4196	> 0.25	0.7750	1.6130	173.3152	Ajusta PDF Gamma
1	1.15	0.5960	0.1353	0.7793	1.3743	161.8739	Ajusta PDF Gamma
2	0.85	0.9402	0.0185	0.7617	3.0927	134.1125	Ajusta PDF Gamma
2	0.95	0.6198	0.1108	0.7608	3.3886	146.2882	Ajusta PDF Gamma
2	1.00	0.5439	0.1687	0.7597	3.7557	159.3044	Ajusta PDF Gamma
2	1.15	0.6211	0.1086	0.7584	4.6281	116.3134	Ajusta PDF Gamma
3	0.85	0.7202	0.0618	0.7577	5.2720	126.4710	Ajusta PDF Gamma
3	0.95	0.7044	0.0691	0.7618	3.0711	194.4100	Ajusta PDF Gamma
3	1.00	0.6095	0.1154	0.7569	6.0990	148.6890	Ajusta PDF Gamma
3	1.15	0.4349	> 0.25	0.7584	4.5806	151.6649	Ajusta PDF Gamma
4	0.85	0.4213	> 0.25	0.7520	63.9444	14.2743	Ajusta PDF Gamma
4	0.95	0.2160	> 0.25	0.7540	13.3896	74.7192	Ajusta PDF Gamma
4	1.00	0.2935	> 0.25	0.7540	12.8446	105.3320	Ajusta PDF Gamma
4	1.15	0.4273	> 0.25	0.7557	7.2692	135.6675	Ajusta PDF Gamma

Anexo 9. Conteo de estados de falla y seguro y sus probabilidades de transición por coeficientes de déficit de la serie simulada y observada.

Serie	λ	# 0/0	# 0/1	# 1/0	# 11	p01	p10
Sintética	0.85	155	309	310	726	0.6659	0.2992
Sintética	0.95	330	337	338	495	0.5052	0.4058
Sintética	1	739	298	299	164	0.2874	0.6458
Sintética	1.15	431	338	339	392	0.4395	0.4637

Serie	λ	# 0/0	# 0/1	# 1/0	# 11	p01	p10
Observada	0.85	5	7	8	30	0.5833	0.2105
Observada	0.95	14	10	11	15	0.4167	0.4231
Observada	1	14	13	13	10	0.4815	0.5652
Observada	1.15	20	13	13	4	0.3939	0.7647