MODELO AUTORREGRESIVO DE PRIMER ORDEN APLICADO A LA PREDICCIÓN ANUAL DE CAUDALES EN LA AMAZONIA PERUANA: CUENCA DEL RÍO MAYO

Santiago Alberto Casas Luna[1,*], Jesús Abel Mejía Marcacuzco[2]

[1] Ingeniero Ambiental, M. Sc. en Ingeniería de Recursos Hídricos, Docente Principal de Universidad Nacional de San Martín (UNSM-T), San Martín – Perú.
[2] Ingeniero Agrícola, Master en Hidrología, PhD. en Ingeniería Civil – Hidráulica y Docente Principal en la Universidad Nacional Agraria La Molina (UNALM), Lima – Perú
* Jr. Chachapoyas Mz.E Lte.01, Moyobamba, San Martín - Perú. centroceap@yahoo.com . Telef: 975492187.

Resumen:

La Amazonía peruana, está en la cabecera de la cuenca del río Amazonas, se caracteriza por presentar diferentes condiciones fisiográficas, hidroclimáticas, ecológicas y ambientales. El comportamiento de las variables hidrológicas en esta zona, son aún poco conocidos debido a la disponibilidad limitada de información. Existe imposibilidad física de contar mediciones en todo los eventos de lluvias ocurridos, por ello los métodos indirectos de estimación de parámetros de flujo, mediante el modelado matemático es una alternativa factible.

La estimación de parámetros de flujos (altura de agua, velocidades y caudales de escurrimiento), son muy importantes para el tratamiento de la mitigación de crecidas, disminución de la carga de sedimentos y agentes contaminantes del escurrimiento, propios de los problemas de la gestión de drenajes pluviales en ámbitos urbanos y rurales.

Los métodos de generación de series temporales se basan en el uso de los “registros históricos” como una muestra de la población total. Muchas series hidrológicas presentan correlación serial; esto significa correlacionar una variable aleatoria X con la misma variable, pero desfasada en un intervalo k de tiempo. El coeficiente de correlación así obtenido se denota por ρx,k y cuando el desfase es k = 1, el proceso será de autorregresivo de primer orden, si k = 2, el proceso es de segundo orden y así sucesivamente. El modelo autorregresivo normal de primer orden (AR1) es usado ampliamente en la generación de caudales anuales.

Palabras claves: Modelo autorregresivo de primer orden, series temporales, generación estocástica de caudales, amazonia peruana, río Mayo.

1. Introducción:

Las cuencas hidrográficas son unidades territoriales donde interactúa simultáneamente el sistema hídrico que produce agua, con el sistema económico, social y cultural; producen bienes y servicios recreativos, agrícolas, pecuarios y forestales. A medida que crece la economía en una región, incrementa la demanda social de una gestión integrada de los recursos hídricos, lo que significa asignar mayor importancia a la evaluación cuantitativa de los procesos que influyen en términos de precisión, magnitud o tendencias del comportamiento de la cantidad y calidad de los flujos de agua (Silva, 2002).

La cuenca del Amazonas es la más grande del planeta, con un área de drenaje de 6 200 000 km2 y un caudal anual promedio de 6 300 km3 de agua que vierte al océano Atlántico (Molinier et al. 1995, Marengo 2006) mencionan que la cuenca del río Amazonas tiene casi el 99% del total de los recursos hídricos existentes en el territorio peruano y un estudio realizado por la UNESCO (2006) considera que la disponibilidad anual neta de agua en esta cuenca es de 2696 mm. La cuenca del río Mayo cubre aproximadamente el 0.14 % del total de la cuenca del Amazonas y el 0.72 % del territorio peruano (Figura 1), se caracteriza por presentar grandes variaciones altitudinales de 200 hasta 4 000 msnm.

Todos los procesos naturales que están en función del tiempo y el espacio, son procesos periódicos y estocásticos, cuyas propiedades son causadas por ciclos astronómicos y la aleatoriedad de los procesos casuales del medio. En consecuencia todos los procesos hidrológicos tienen características estocásticas o una combinación de procesos determinísticas y estocásticas (Yevjevich, 1972).

El problema frecuente es la insuficiencia de datos, sean de precipitación o, de manera más común, de caudales, en la mayoría de casos, se asume que el futuro es estadísticamente similar al pasado y es sostenido por la hidrología estocástica (García, 2010).

Los modelos estocásticos o de series temporales, carecen de bases físicas y expresan en términos de probabilidad el resultado de procesos altamente aleatorios. El origen de la estocasticidad en los recursos hídricos es doble: La distribución de la lluvia es un proceso aleatorio (prácticamente puro) y los diferentes factores que afectan a la propagación del agua en la superficie terrestre (conductividad, flujo de agua, cubierta vegetal, y otros) le confieren también cierta aleatoriedad (Marco, 1993).

Actualmente existen muchos modelos para la simulación de caudales, entre los que están: redes neuronales artificiales, análisis espectral singular, modelos de ajuste no lineal por partes, modelos de ondas, modelos adaptativos de regresión múltiple, modelos estocásticos AR(p), MA(q), ARMA(p,q), ARIMA(p,i,q)); a partir de estos modelos se derivan modelos lineales y no lineales para la predicción de caudales. La alta no linealidad de los procesos físicos en una cuenca y que generan los caudales, requiere de nuevos enfoques tanto en los modelos físicos como en los matemáticos (Cadavid, J. et al. 2013).

El objetivo general de este trabajo es desarrollar un modelo autorregresivo de primer orden aplicado a la predicción anual (100 años) de caudales del río Mayo, mediante el análisis de datos históricos, generación de series temporales de caudales con distribución comparada normal, log normal y gamma; con el fin de lograr una predicción de caudales confiables.

2. Metodología:

2.1 Área de estudio

La cuenca del río Mayo se ubica al nor oeste de la cuenca del río amazonas (Figura 1), geopolíticamente entre los departamentos de Amazonas, San Martín y Loreto y se extiende por las provincias de Jumbilla (Amazonas); Rioja, Moyobamba, Lamas y San Martín (San Martín) y Alto Amazonas (Loreto); el área de estudio comprende aguas arriba a la estación Shanao que abarca una superficie de 8 372 km². Su altitud máxima es de 4000 msnm y la mínima 268 msnm en la estación hidrométrica Shanao. El curso principal se denomina río Mayo y su trayecto es de nor-oeste a sur-este y es uno de los principales afluentes del río Huallaga y este a su vez principal afluente del río Amazonas.

Figura 1. Localización del área de estudio y estaciones hidrológicas

La serie disponible para el análisis corresponde a caudales mensuales medidos por SENAMHI (Servicio Nacional de Meteorología e Hidrología) en la estación de Shanao – rio Mayo del Departamento de San Martín desde Enero del año 1983 a Diciembre del año 2013 (31 años).

Año	*Q (m^3/s)*	*Año*	*Q (m^3/s)*	*Año*	*Q (m^3/s)*	*Año*	*Q (m^3/s)*
1983	369,99	1991	448,23	1999	384,77	2007	432,61
1984	373,59	1992	456,08	2000	372,87	2008	426,98
1985	407,75	1993	444,82	2001	362,69	2009	428,78
1986	412,99	1994	425,99	2002	357,52	2010	431,45
1987	401,35	1995	413,25	2003	377,96	2011	434,88
1988	423,51	1996	397,12	2004	370,43	2012	444,07
1989	438,33	1997	397,13	2005	374,94	2013	462,40
1990	451,87	1998	366,03	2006	401,93		

Tabla 1. Caudales Observados en la Estación Shanao – Río Mayo

Debido a que las series de caudales, en general presentan correlación serial, significa que vamos a correlacionar una variable aleatoria X con la misma variable, pero desfasada en un intervalo k de tiempo. El coeficiente de correlación así obtenido se denota por ρx,k; y cuando el desfase es k = 1 el proceso autorregresivo es de primer orden.

Los métodos de generación de series temporales autorregresivos (estocásticos), se basan en el uso de los “registros históricos” como una muestra de la población total. Mientras que los métodos convencionales (determinísticos) consideran a los registros como la “población total”. De ello, se deduce que el diseño está basado en estimaciones de lo que “podría haber pasado”, en vez de “lo que ha pasado”.

De esta manera, para desarrollar el modelo AR(p), conocido también como el modelo de Markov de primer orden, se sostiene en la siguiente ecuación [1]:

$$X_{i+1} = \mu_X + \rho_{X,1}(Xi - \mu_X) + \varepsilon_{i+1} \qquad [1]$$

Xi es el valor del proceso en el tiempo i, μ_x es la media de *X,* $\rho_{X,1}$ es el coeficiente de correlación serial de orden 1 y ε_{i+1} es un componente aleatorio con la media, E(ε) = 0 y variancia σ^2. Se asume que ε_{i+1} es independiente de Xi. La variancia de X y ε se relacionan mediante las siguientes ecuaciones [2], [3] y [4]:

$$Var(X_{i+1}) = \sigma^2_X = E[\mu_X] + \rho_{X,1}(Xi - \mu_X) + \varepsilon_{i+1}]^2 - E^2(X_{i+1}) = \rho^2_{X,1}\,\sigma^2_X + \sigma^2_\varepsilon \qquad [2]$$

$$\sigma^2_\varepsilon = \sigma^2_X\, 1 - \rho^2_{X,1} \qquad [3]$$

$$\sigma_\varepsilon = \sigma_X\, [\,1 + \rho^2_{X,1}]^{1/2} \qquad [4]$$

Si $\rho_{X,1} = 0$, la ecuación [1] se reduce a $X_{i+1} = \mu_X + t_{i+1}\sigma_x$ y si $\rho_{X,1} = 0$, $X_{I+1} = X_i$ y el modelo se hace completamente determinístico quedando el valor de X_{i+1} completamente determinando por X_i.

Para el proceso autorregresivo (Markov) de primer orden, el coeficiente de correlación serial con desfase K está dado por:

$$\rho_{x,k,} = \rho^k_{X,1} \quad [\,5\,]$$

Por lo tanto el correlograma decae exponencialmente desde $\rho_{X,0} = 1$ hasta $\rho_{X,\infty} = 0$. Si se observa que el correlograma tiene esta propiedad el modelo de Markov puede considerarse como el apropiado para la generación de datos.

La metodología para analizar, si las series observadas son o no estacionarias, consiste en obtener, además de los gráficos de las series, los correlogramas simples o función de autocorrelación simple (FACS) y función de autocorrelación parcial (FACP). Los coeficientes de autocorrelación proporcionan información sobre la relación lineal entre los residuos del modelo separados por k unidades temporales. Es decir, indican el grado de correlación entre cada valor del residuo y los desplazados 1, 2, ... h periodos

La verificación del modelo se realiza observando la función autorregresivo simple (FACS) de los residuos, que es significativamente nula para cualquier retardo. La esencia básica para el análisis estocástico es que el proceso sea estacionario, es decir, que las propiedades estadísticas del proceso no varían en el tiempo. Así, las propiedades de los registros históricos se pueden utilizar para derivar series sintéticas largas.

2.2. Resultados y Discusión

Los cálculos se realizaron mediante un código computacional en MATLAB, obteniéndose los siguientes resultados:

Medidas de tendencia central y dispersión

La media, explica a la medida de tendencia central de los datos de caudales generados del río Mayo: xm= 408,9945.

La desviación estándar explica a la medida de dispersión de los datos de caudales generados del río Mayo, respecto a la media: sx = 31.9438

Los coeficientes de autocorrelación:

Mediante el modelo autorregresivo, se describe un proceso en que las observaciones actuales se hallan determinadas o influidas por las observaciones anteriores.

Considerando que la serie anual de caudales del río Mayo es de 31 datos, los coeficientes de autocorrelación obtenidos, hasta k = 16 (retardos), se presentan en la Tabla 2.

k	R	k	R	k	r	k	r
1	1,0000	5	0,3161	9	-0,7193	13	-0,7973
2	0,8824	6	0,0014	10	-0,8569	14	-0,5580
3	0,6996	7	-0,2532	11	-0,9017	15	-0,2697
4	0,5306	8	-0,5032	12	-0,8947	16	-0,0859

Tabla 2. Coeficientes de autocorrelación, hasta k = 16

Sin embargo, a partir de los casos estudiados (140 ríos de todo el mundo) por Yevjevich (1963), se infiere que hay una gran variabilidad entre los parámetros estadísticos de series de caudales anuales para diferentes ríos. Donde el parámetro más difícil de estimar es el coeficiente autorregresivo, necesitándose una muestra superior a 40 observaciones para que la probabilidad de confianza del intervalo sea al menos del 75%, además de la media y varianza, para que los errores de las estimaciones no sean muy elevados y la aplicación de la metodología sea confiable en los resultados.

Correlograma e identificación del modelo

Para interpretar mejor la forma de los correlogramas, hay que tener en cuenta que, entre los factores que influyen en el régimen fluvial el clima es el más importante, especialmente las precipitaciones y su distribución.

En el presente estudio, a partir del correlograma (figura 2), el modelo AR1 se define como estacionaria, debido a que los coeficientes de autorregresión con valores positivos y negativos se aproximan finalmente al valor cero.

Del análisis del correlograma (figura 3) con un límite de confianza al 95 %, estimado a partir de los datos de la serie anual de caudales del río Mayo, con k = 16 (retardos) y coeficientes de autocorrelación que oscilan entre 1.00 a -0.9017, se reafirma que el modelo AR1 es estacionaria.

Entonces, algunas propiedades de las series temporales hidrológicas pueden ser investigadas en un dominio temporal a través del análisis de correlogramas. Cuando se hayan detectado ciertas “tendencias” (trends en Inglés), estas deben ser quitadas de las series originales y, la serie de los residuales es la que se debe examinar, el interés se centra en la distribución de probabilidad de los elementos de la serie de residuales.

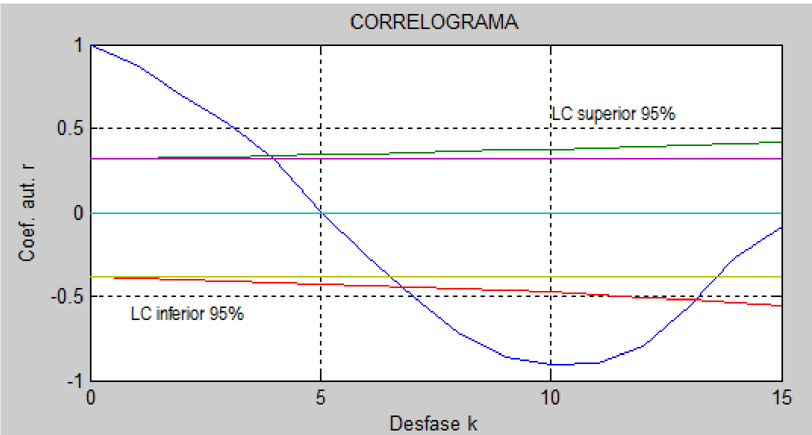

Figura 2. Correlograma a partir de la serie histórica de caudales anuales del río Mayo

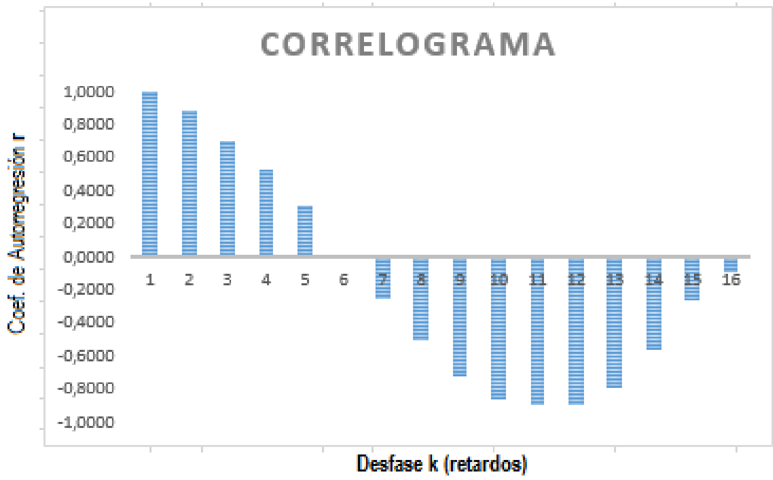

Figura 3. Correlograma del modelo AR(1) estacionaria, río Mayo

Generación de series anual de caudales

Es factible conocer (estimar) el comportamiento futuro de los caudales anuales del río Mayo en diferentes series de tiempo, en este caso los caudales futuros a 100 años en el punto de aforo de la estación Shanao, con la finalidad de simular procesos hidrológicos importantes para la planificación del uso o aprovechamiento de la disponibilidad anual del agua, en el marco de la gestión integrada de los recursos hídricos. En la tabla 3, tenemos los caudales generados para 100 años, a partir de la serie histórica de caudales anuales de 31 años, del río Mayo en el punto de aforo de la estación Shanao.

Año	Q (m^3/s)	Año	Q (m^3/s)	Año	Q (m^3/s)	Año	Q (m^3/s)	Año	Q (m^3/s)	Año	Q (m^3/s)
1	369,99	18	432,31	35	375,27	52	411,46	69	419,42	86	433,46
2	383,31	19	397,65	36	361,14	53	421,16	70	422,83	87	425,84
3	353,38	20	391,40	37	362,96	54	421,01	71	417,69	88	430,30
4	325,06	21	374,37	38	346,90	55	432,84	72	400,78	89	412,22
5	336,13	22	372,69	39	353,89	56	434,89	73	397,47	90	440,07
6	330,45	23	386,71	40	351,94	57	420,06	74	397,53	91	450,55
7	345,87	24	401,74	41	391,39	58	391,62	75	376,79	92	457,50
8	363,47	25	387,34	42	410,57	59	421,60	76	383,47	93	438,63
9	381,72	26	382,81	43	372,85	60	411,03	77	374,11	94	439,95
10	374,54	27	387,95	44	383,74	61	412,35	78	376,80	95	427,92
11	385,34	28	386,04	45	365,69	62	420,42	79	385,64	96	421,01
12	389,64	29	393,27	46	366,95	63	420,78	80	374,79	97	411,03
13	404,33	30	401,13	47	374,37	64	405,80	81	374,48	98	395,37
14	412,94	31	388,08	48	389,68	65	399,14	82	383,80	99	383,32
15	425,97	32	387,88	49	387,85	66	398,42	83	359,17	100	383,18
16	421,99	33	358,32	50	414,03	67	421,90	84	380,60		
17	418,25	34	381,49	51	406,21	68	407,44	85	420,38		

Tabla 3. Caudales anuales generados para 100 años río Mayo – Estación Shanao

Serie de caudales anuales históricos y generados con distribución normal

En la figura 4, se observa que la generación de serie anual de caudales generados a 100 con distribución normal, subestima los primeros 10 años y sobrestima (del año 11 al 31), respecto a la serie histórica en el punto de aforo Shanao del río Mayo. Por lo tanto, significa que la distribución normal de la serie temporal, tiene propiedades estadísticas variables, en el proceso de generación de caudales a paso anual para 100 años con distribución normal, del modelo AR.

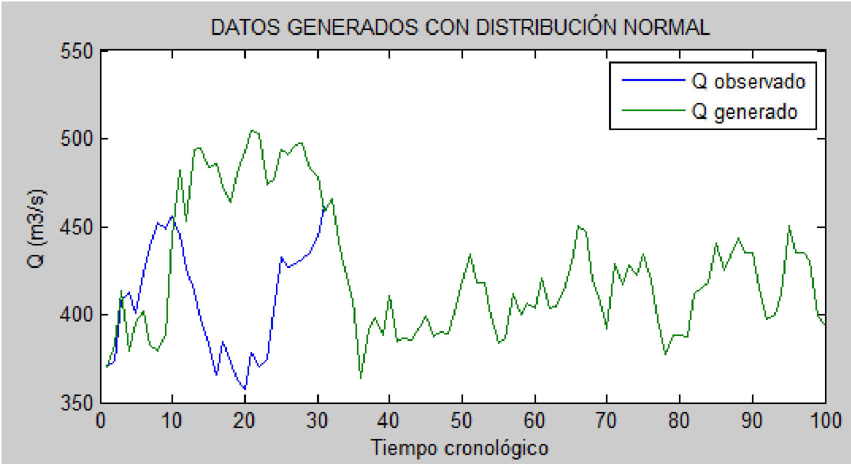

Figura 4. Serie anual de caudales generados con distribución normal, río Mayo – Estación Shanao

Serie de caudales anuales históricos y generados con distribución log normal

En la figura 5, se observa que la generación de serie anual de caudales generados a 100 con distribución log normal, subestima significativamente los primeros 15 años y entre los año 26 al 31, sobrestima entre el año 15 al 26; respecto a la serie histórica en el punto de aforo Shanao del río Mayo. Por lo tanto, significa que la distribución de la serie temporal, tiene propiedades estadísticas variables, en el proceso de generación de caudales a paso anual para 100 años con distribución normal, del modelo AR1 y requiere mejor ajuste.

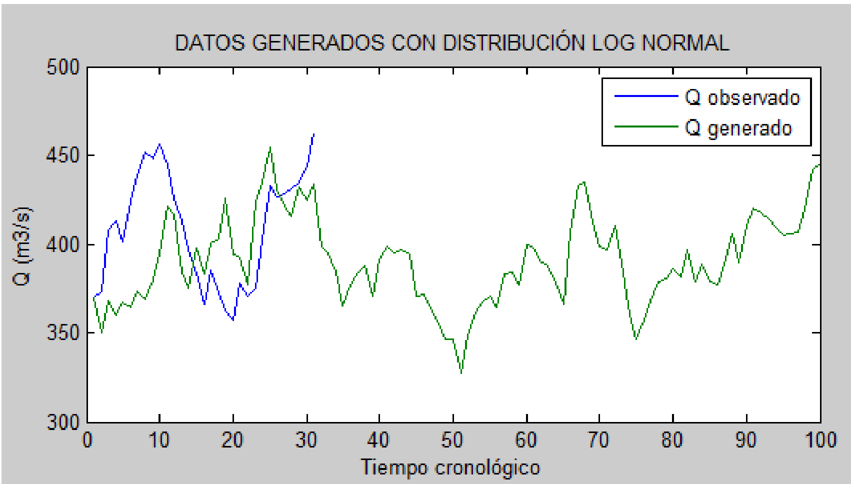

Figura 5. Serie anual de caudales generados con distribución log normal, río Mayo – Estación Shanao

Serie de caudales anuales históricos y generados con distribución gamma

En la figura 6, se observa que la generación de serie anual de caudales generados a 100 con distribución log normal, subestima significativamente los primeros 12 años y sobrestima entre los años 12 al 31; respecto a la serie histórica en el punto de aforo Shanao del río Mayo. Por lo tanto, significa que la distribución gamma de la serie temporal, tiene propiedades estadísticas variables, en el proceso de generación de caudales a paso anual para 100 años con distribución normal, del modelo AR1.

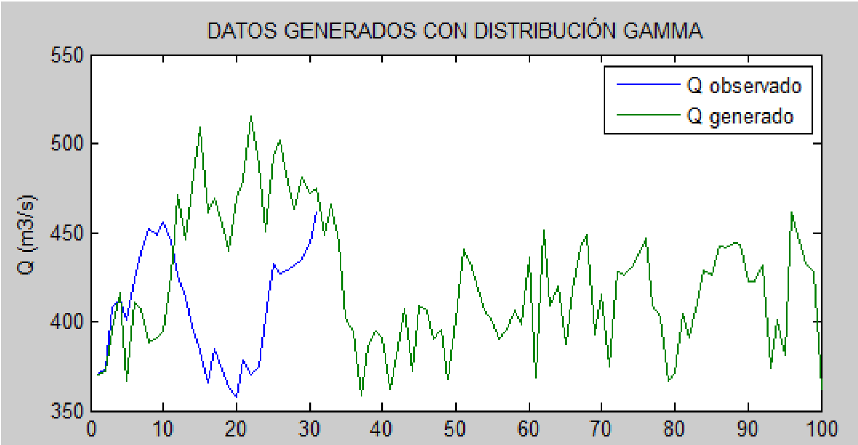

Figura 6. Serie anual de caudales generados con distribución gamma,
río Mayo – Estación Shanao

En la figura 4, 5 y 6 se presenta la serie anual de caudales generados a partir de la serie histórica de caudales anuales (1983 – 2013) con distribución normal, log normal y gamma para los próximos 100 años. Del análisis de las distribuciones de caudales generados podemos observar a la distribución que más se ajusta a la serie histórica de caudales anuales estudiados. Es decir las series de caudales generados para 100 años, deben mostrar propiedades estadísticas similares a aquéllas de la serie histórica.

3. Conclusiones:

El desarrollo del modelo autorregresivos (AR1) para la generación de series anuales de caudales para 100 años, a partir de la serie histórica (1983-2013) en el punto de aforo de Shanao del río Mayo, con distribución normal, log normal y gamma, tienen propiedades estadísticas variables; por lo que, para la fase de simulación hidrológica, se requieren de un mejor ajuste mediante procesos de calibración y validación. Sin embargo, en referencia al punto de aforo Shanao, la generación de serie anual de caudales con distribución log normal, es el que representa mejor acercamiento al comportamiento de la serie histórica; por lo tanto, ofrece mayor confianza para la predicción anual y descripción del comportamiento para 100 años del río Mayo.

Referencias

Cadavid, J. & Carvajal, L. (2013). Modelo autorregresivo bilineal aplicado a la predicción mensual de caudales en Colombia. Medellín, Colombia.

Fiering, M. & Jackson, B. (1971). Synthetic Streamflows. Colorado, EE.UU.

García, F. P. (2010). Modelación hidrológica estocástica: Desarrollo de un modelo de generación sintética de series temporales. Santa Cruz, Bolivia.

Marco, J. (1993). Hidrología estocástica y planteamiento hidráulico en conceptos y métodos para la planificación hidrológica. Barcelona, España.

Marengo, J. A. (2006). On the hydrological cycle of the amazon basin: a historical review and current state of the arty. São Paulo, Brazil.

Molinier, M., Guyot, J. L., De Olivera, E. & Guimarães, V. (1995). Les régimes hydrologiques de l´Amazone et de ses affluents. Brasilia DF, Brésil.

Silva, O. (2010). Evaluación del modelo hidrológico SWAT en la cuenca media del río Pao en Venezuela.

Yevjevich, V. (1972). Probability and statistics in hydrology. Fort Collins, Colorado, USA.