

Stochastic modeling of annual average flows basin in Santa river, Peru

Abelardo M. Díaz Salas^a, Edilberto Guevara^{*,**,b}

^aEscuela de Post Grado de la Universidad Nacional Agraria La Molina. Doctorado en Ingeniería de Recursos Hídricos, Lima, Perú

^bInstituto Hidrología Aplicada y Medio Ambiente (IHAMA), Lima, Perú

Abstract.-

The interpretation of temporal and spatial behavior average unloading annual were done through stochastic models of stationary series, where we have found the first orden autoregressive model AR(1) convenient. The regional interpretation for temporal and spatial behavior of annual average unloading were done through a regional model which is in accordance with the area of the basin, percentage of glacier area, the annual regional parameter (φ_1) and regional model of the standard deviation of error S_e . With these regional models we can generate volumes of water or unloading annuals averages in any part of the basin knowing the area of the basin, the (φ_1) annual regional parameter and regional model of the standard deviation of error. To verify the validity of the modeling we have compared average statistics and variance of the historical series and of the generated series, where we have found acceptable results; that is to say, that we can infer that these data come from the same population.

Keywords: stochastic modeling hydrological series modeling; synthetic generation of volume of water

Modelación estocástica de los caudales medios anuales en la cuenca del río Santa, Perú

Resumen.-

La interpretación del comportamiento temporal y espacial de las descargas medias anuales se ha realizado a través de modelos estocásticos de series estacionarias, donde se ha encontrado al modelo autorregresivo de primer orden AR(1) como adecuado. La interpretación regional de las descargas medias anuales se ha realizado mediante un modelo que está en función del área de la cuenca, porcentaje del área glaciar, del parámetro regional anual (φ_1) y la desviación estándar del error S_e . Mediante estos modelos regionales se pueden generar los caudales o descargas medias anuales en cualquier punto de la cuenca, conociendo el área de la cuenca, el parámetro regional anual (φ_1) y modelo regional de la desviación estándar del error S_e . Para verificar la validación del modelamiento se ha comparado las estadísticas media y variancia de las series históricas y de las series generadas, donde se ha encontrado resultados aceptables, es decir se puede inferir que los datos provienen de una misma población.

Palabras clave: modelamiento estocástico de series hidrológicas; generación sintética de caudales medios

Recibido: enero 2016

Aceptado: julio 2016

1. Introducción

En la cuenca del río Santa existen varios proyectos que requieren estimar la oferta de agua como son los caudales medios anuales, en lugares donde no existen datos hidrológicos, la generación de estos caudales se puede realizar con los modelos estocásticos regionales.

El objetivo del presente trabajo es determinar modelos estocásticos regionales que permitan

* Autor para correspondencia

** Dirección permanente: Escuela de Ingeniería Civil, Facultad de Ingeniería, Universidad de Carabobo. Venezuela
Correo-e: eguevara@ana.gob.pe (Edilberto Guevara)

generar descargas medias anuales en sitios no aforados en la cuenca del río Santa.

Los Objetivos específicos son: determinar el área de la cuenca, el porcentaje del área glaciar de cada estación de aforo y regionalizar los parámetros de los modelos estocásticos ARMA(p, q) para las descargas medias anuales.

Para el modelamiento estocástico se ha trabajado en 9 ríos de la cuenca del río Santa que son: río Yanayacu, río Negro, río Quillcay, río Marcará, río Llanganuco, río Parón, río Colas, río Los cedros y río Quitaraca. Cada estación de aforo para el modelamiento tiene 41 años de registro de caudales comprendidos entre el periodo de 1968 a 2008.

El modelo regional adecuado para las descargas medias anuales es el modelo estocástico de autorregresivo de primer orden ARMA(1,0). El modelo regional para estimar las descargas medias anuales está en función del área de la cuenca, porcentaje del área glaciar, del parámetro regional (φ_1) y la desviación estándar del error (está en función del área de la cuenca).

Estimar la oferta anual de agua, a través del tiempo en sitios no aforados en el ámbito de la Cordillera Blanca del río Santa actualmente no es posible porque no existen modelos estocásticos validados o contrastados con los datos observados. Los proyectos para captar el agua no están ubicadas en las estaciones de aforo y por lo tanto es necesario determinar modelos de series de tiempo.

Mediante el modelamiento estocástico utilizando datos de estaciones de aforo, se generan series sintéticas en sitios sin aforo es decir en cualquier punto de la cuenca, para lo cual se debe tener como dato el área de la cuenca, porcentaje del área glaciar, parámetro regional (φ_1) y la desviación estándar del error.

1.1. Antecedentes

Existen varios estudios de cuencas de ríos por series temporales, entre los cuales destacan los trabajos de Guevara *et al* [1], el de Wu y Chau [2], el de Sang [3], y el de Fry *et al* [4].

2. Fundamento teórico

2.1. Modelo estocástico

Custodio y Llamas [5] indican que los modelos estocásticos son adecuados para describir las series cronológicas o series de tiempo hidrológico. Las series cronológicas se dividen en dos categorías: determinísticas y estocásticas. Las series determinísticas son aquellas cuyo valor en cualquier instante se puede determinar con certeza y estas series pueden ser periódicas, semiperiódicas, o presentar tendencias o saltos. Las series estocásticas pueden ser estacionarias o no según sus parámetros (media, varianza, etc.) sean independientes o no a través del tiempo.

2.2. Series de tiempo o series cronológicas

Pindyck y Rubinfeld [6] indican que cada valor de y_1, y_2, \dots, y_t en la serie es extraído al azar de una distribución de probabilidad. Al modelar dicho proceso, se intenta describir las características de su aleatoriedad, esto ayudaría a inferir algo sobre las probabilidades asociadas con los valores futuros alternativos de la serie. Para generalizar se puede decir que la serie observada y_1, \dots, y_t es extraída de un conjunto de variables aleatorias distribuidas en forma conjunta. Si de alguna manera se pudiera especificar la función de distribución de probabilidad de la serie entonces se podría determinar la probabilidad del resultado futuro. Sin embargo la especificación completa de la función de la distribución de probabilidad para la serie de tiempo por lo general es imposible. Por lo tanto es usual construir un modelo simplificado de las series de tiempo que explica su aleatoriedad de manera que sea útil para pronosticar los valores futuros que son los modelos estocásticos (modelos autorregresivos, modelos de promedio móvil, etc).

2.3. Análisis de las series temporales

Hernández [7] indica que históricamente tres son los tipos de análisis o procedimientos metodológicos para la evaluación o estudio de series temporales como son: los métodos de descomposición, método de análisis causal y el método de Box-Jenkins.

El método de descomposición, consiste en dividir la serie temporal en sus componentes como son: tendencia, ciclo, componente estacional y componente irregular (aleatorio). El método de análisis causal, consiste en explicar la evolución de la serie temporal mediante una ecuación que relaciona la variable en estudio con otras variables que inciden sobre ella y el Método de Box–Jenkins, consiste en explicar evolución de la serie temporal (variable temporal) en función del comportamiento en el pasado de la misma serie temporal. Conocida la estructura de dependencia entre valores sucesivos de la serie mediante modelos integrados autorregresivos de medias móviles (ARIMA) es posible predecir la evolución futura de la serie temporal a través de su pasado.

2.3.1. Series de tiempo estacionarios

Gujarati y Porter [8] indican que un proceso estocástico es estacionario en el *sentido dévil* si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende sólo de la distancia o rezago entre estos dos periodos y no del tiempo en el cual se calculó la covarianza. 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. Esta propiedad estacionaria se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo. Una serie temporal dévilmente es estacionaria si

$$E[X_t] = \mu \quad (\text{constante}) \quad \forall t; \quad (1)$$

$$\text{VAR}[X_t] = \sigma^2 \quad (\text{constante}) \quad \forall t; \quad (2)$$

$$\text{COV}[X_t, X_{t+k}] = \gamma_k \quad (\text{independiente de } t, \forall k. \quad (3)$$

2.3.2. Prueba de estacionaridad

Peña [9] muestra la pruebas de estacionariedad de Dickey–Fuller aumentado como la prueba estándar para verificar la estacionariedad de una serie.

2.3.3. Modelos ARMA(p, q)

Existe una amplia literatura de la estimación de modelos para series estacionarias, por ejemplo Pindyck y Rubinfeld [6], Peña [9], Gujarati y Porter [8], y Court y Rengifo [10]. Donde están definidos los modelos autorregresivos AR(p), los modelos media móvil MA(q) y los modelos ARMA(p, q).

2.3.4. Estimación y validación del modelo

Para determinar el modelo adecuado para explicar el comportamiento de una serie de tiempo, Gujarati y Porter [8] y Peña [9] indican una metodología consistente en cuatro pasos: identificación, estimación, examen de diagnóstico y pronóstico.

2.3.5. Análisis de las series de tiempo hidrológico

Para el análisis de series hidrológicas es importante evaluar las propiedades estadísticas de las series de descargas medias anuales. Además, es importante identificar y evaluar las tendencias, los saltos, la estacionalidad y la no normalidad. El análisis de las series de tiempo hidrológico se ha convertido en una herramienta importante en la hidrología, es utilizado para la construcción de modelos matemáticos, para generar registros sintéticos de hidrología, para pronosticar los eventos hidrológicos, para detectar tendencias, saltos, para llenar datos incompletos y para extender o ampliar los registros de datos.

2.4. Componentes de la estructura de la serie de tiempo hidrológico

Una serie de tiempo hidrológico puede ser dividido o descompuesto en sus componentes, como son: tendencias, cambios o saltos, estacionalidad, etc.

2.4.1. Saltos y tendencias

En general, los factores naturales y provocados por el hombre pueden producir tendencias graduales y cambios saltos instantáneos series hidrológicas.

Maidment [11] indica que el calentamiento global y los cambios climáticos están creando conciencia en los hidrólogos de la existencia de tendencias y saltos en las series de tiempo

hidrológico. Para fines de pronóstico se deben tener series hidrológicas libres de tendencia y saltos.

2.5. Diagnóstico del modelo y pronóstico

Court y Rengifo [10] indica que para verificar si el modelo seleccionado es adecuado, existen dos métodos: el primero consiste en crear un modelo con más rezagos que los identificados en el paso previo y verificar que los rezagos extra no son estadísticamente significativos. El segundo método consiste en analizar los residuos del modelo seleccionado. Si el modelo seleccionado captura adecuadamente la dinámica de la variable, los errores deberían ser ruidos blancos, es decir no autocorrelacionados.

Tanto Peña [9] como González [12] indican que la diagnóstico del modelo requiere comprobar que las hipótesis básicas realizadas a los residuos sean ciertas. Los residuos deben tener

1. media marginal, igual a cero;
2. variancia marginal constante;
3. no debe existir correlación para cualquier retardo.

Adicionalmente se acostumbra pedir que los residuos tengan distribución normal.

Estos contrastes son

1. Para comprobar si la media es cero, la clásica prueba de *t*-Student es usada con hipótesis nula $H_0 : \mu = 0$.
2. Para comprobar si la varianza es constante, la prueba *F* de Fisher es usada para comparar la varianza de dos intervalos *I* y *J* de los residuos, con hipótesis nula $H_0 : \sigma_I^2 = \sigma_J^2$.
3. Ausencia de correlación serial

Si los residuos se comportaran como un ruido blanco, los coeficientes de la FAC y FACP muestrales deben ser prácticamente nulos para todos los retardos. Para comprobarlo, se utiliza el estadístico de contraste de Ljung-Box dado por la Ecuación 4.

$$Q(M) = T(T+2) \frac{\sum_{k=1}^M \hat{\rho}_k^2(a)}{T-k} \chi^2(M-p-q), \quad (4)$$

donde *M* es el número de rezagos, *T* tamaño de la muestra.

Se rechaza la hipótesis nula de ausencia de correlación serial, para un nivel de significación del 5%, para valores grandes del estadístico, es decir, si:

$$Q(M) > \chi_{0,05}^2(M-p-q).$$

En este caso, existiría correlación serial en los residuos del modelo por lo que se concluye que el modelo no ha sido capaz de reproducir el patrón de comportamiento sistemático de la serie y habría que reformularlo.

Esta prueba es necesaria porque garantiza que la incorrelación implica la independencia y que no se está dejando información por modelizar (ver Peña [9]).

4. Pruebas complementarias, las siguientes pruebas complementan el diagnóstico del modelo.

- a) Prueba de normalidad, la clásica prueba de normalidad suele ser complementada por las pruebas de kurtosis y sesgo indicada por Sveinsson *et al* [13].
- b) Prueba de rachas arriba y abajo de la mediana.
- c) Prueba de rachas de subidas y bajadas.

Las pruebas de rachas son test no paramétricos que están en cualquier libro clásico de estadística.

3. Metodología

La modelación de las series de descarga anuales se ha realizado siguiendo el procedimiento indicado en la Figura 1.

3.1. Recopilación de datos

Los datos recopilados para el modelamiento estocástico, son las descargas medias anuales y los parámetros físicos de la cuenca como: área de la cuenca, y el porcentaje del área glaciar. Estos datos se han obtenido de ANA [14].

Los registros de descargas medias anuales que se han recopilado están comprendidas entre los años de 1968 al 2008 (41 años) en las

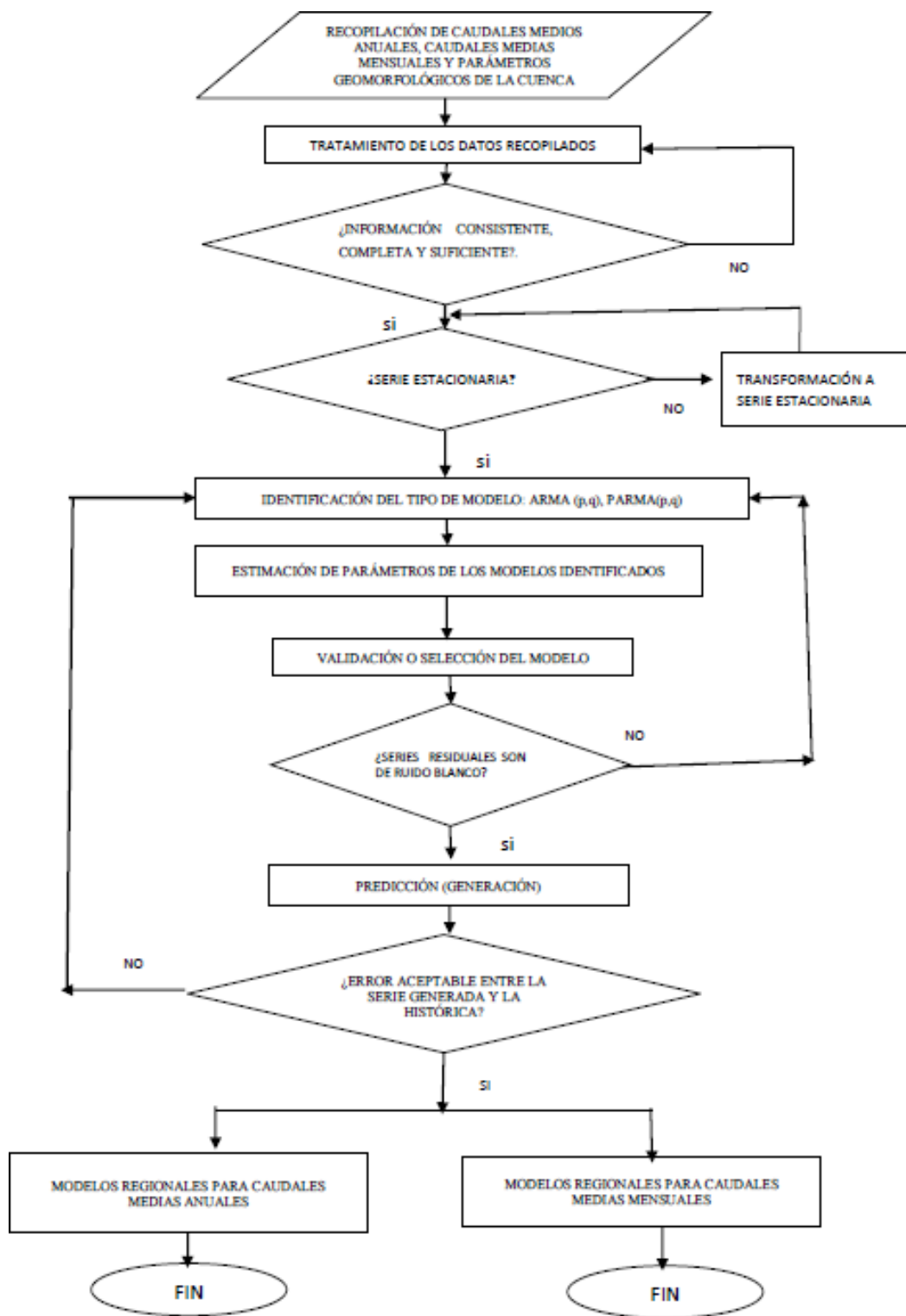


Figura 1: Diagrama de flujo para estimación de modelos de series hidrológicas.

9 estaciones estudiadas (Querococha, Olleros, recopilado de Suarez [15]. Quillcay, Chancos, Llanganuco, Parón, Colcas, Los Cedros y Quitaracsa). Estos datos se han

3.2. Tratamiento de datos

El tratamiento de las descargas de las cuencas en estudio ha consistido en el análisis de consistencia de las descargas de los ríos. Teniendo una información consiste el siguiente paso es completar los datos faltantes y extender los datos de descargas de estaciones hidrológicas de registro corto a fin de tener información hidrológica consistente, completa y suficiente.

3.3. Prueba de estacionaridad

Los modelos ARMA(p, q) se aplican para series estacionarias. Para verificar la estacionaridad de las descargas medias anuales en las 9 estaciones estudiadas se ha utilizado la prueba de raíces unitarias, la descripción de esta prueba se encuentra en Peña [9].

3.4. Identificación del tipo de modelo estocástico

Teniendo las series hidrológicas consistentes, completas, suficientes (extendidas) y estacionarias, el siguiente paso es buscar el o los modelos adecuados para las series hidrológicas de caudales medios anuales. Los modelos usados frecuentemente en la modelación de las series hidrológicas no periódicas son: modelos autorregresivos simbolizados como modelos AR (p), los modelos de media móvil simbolizados como modelos M(q) y los modelos autorregresivos y de media móvil simbolizados como modelos ARMA(p, q).

En esta etapa el objetivo no es seleccionar un modelo, sino identificar un conjunto de modelos posibles que sean compatibles con el gráfico de las sus funciones de autocorrelación simple y parcial. En la práctica la mayoría de series reales pueden aproximarse bien con los modelos ARMA(p, q) con p y q menores que tres para series no estacionales, y con P y Q menores que dos, para series estacionales. La identificación de los posibles modelos requiere detectar la estacionaridad (estabilidad) o la no estacionaridad de la serie de tiempo, si sucede el último caso (series no estacionaria), es necesario realizar transformaciones para obtener series estacionarias, que se pueden modelar mediante los modelos ARMA (Peña, [9]).

3.5. Estimación de los parámetros de los modelos identificados

Después de identificar el tipo y orden del modelo, se estiman los parámetros del modelo. Los parámetros de los modelos se estiman con las ecuaciones dadas por Sveinsson *et al* [13]. Los parámetros a estimar son μ , (φ), (θ) y la variancia del ruido blanco ($\widehat{\sigma}_\varepsilon^2$). Los parámetros depende del tipo de modelo ARMA(p, q).

3.6. Validación o selección del modelo adecuado

Del conjunto de modelos identificados M^1, \dots, M^m , y después de estimar sus parámetros se selecciona al modelo que explica mejor al comportamiento de la serie observada. La validación del modelo se efectúa con los criterios de Akaike, y Schwars.

Sveinsson *et al* [13] indican que la validación del modelo se realiza comparando los criterios de Akaike, y Schwars modificados, y se usa más el criterio de información de Schwarz modificado.

3.7. Examen de diagnosis

Para verificar si el modelo seleccionado con el criterio establecido es adecuado se debe analizar los residuos de este modelo. Si el modelo seleccionado captura adecuadamente la dinámica de la variable, los errores deberían ser ruidos blancos.

Para comprobar que los residuos son ruidos blancos debe cumplir con las condiciones del apartado 2.5.

3.8. Predicción o generación

Después de validar el tipo modelo y comprobar que los residuos son ruidos blancos se genera la serie hidrológica, reemplazando los parámetros estimados y generando los ruidos blancos que deben tener media cero y varianza constante.

3.9. Estimación de los parámetros (μ), (φ), zeta (θ)

Los parámetros (φ) zeta (θ) se estiman con las ecuaciones según el modelo validado usando las ecuaciones dada por Sveinsson *et al* [13].

3.10. Generación de datos

Definido el modelo adecuado se reemplazan los parámetros μ , (φ) , zeta (θ) y ε_τ en las ecuaciones correspondientes de ARMA(p, q), y se obtienen los valores de datos generados.

3.11. Verificación de la bondad de ajuste del modelo estocástico

Existen muchos métodos y criterios para verificar la bondad de ajuste del modelo, el criterio principal consiste en comparar las características estadísticas de las series generadas con las series históricas de la misma longitud. Si el modelo es adecuado las estadísticas que usualmente deben de preservarse son: la media y la desviación estándar, la estructura de dependencia, la distribución de probabilidades y la persistencia de los años húmedos y años secos.

Para la verificación de la validez del modelo estocástico es necesario la realización de pruebas estadísticas en los parámetros más importantes como la media y la desviación estándar, usando la prueba de medias y la prueba de homogeneidad de variancias entre los datos de las series históricas y de las series sintéticas generadas.

3.12. Regionalización de los modelos estocásticos de caudales medios anuales

El tipo y orden del modelo estocástico se define siguiendo la secuencia establecida en la Figura 1.

3.13. Modelo estocástico regional para caudales medios anuales en la cuenca del río santa

Después de definir el modelo adecuado que puede ser modelo estocástico autorregresivo AR(p), modelo estocástico de media móvil MA(q) o modelo estocástico autorregresivo y de media móvil ARMA(p, q), se busca un modelo regional para los parámetros media μ , (φ) zeta (θ) y la desviación estándar del error (S_e). Estos parámetros regionales permitirá generar caudales medios anuales en puntos no aforados de la cuenca del río Santa.

4. Análisis y discusión de resultados

El modelamiento estocástico de las descargas medias anuales, se ha aplicado a 9 ríos de la cuenca del río Santa.

4.1. Recopilación de datos

El periodo de registro de las descargas medias anuales es de 41 años en 9 ríos de la cuenca río Santa lo cual se muestra en la Tabla ???. También se han recopilado datos de parámetros físicos de las cuencas. Las características geomorfológicas de los ríos afluentes de la cuenca del río Santa están recopilados de ANA [14].

4.2. Descripción de la cuenca del río Santa

Políticamente la cuenca del río Santa está comprendida en los departamentos de Ancash (provincias de Recuay, Huaraz, Carhuaz, Yungay, Huaylas, Corongo, Pallasca y Santa) y la Libertad (provincias de Virú y Santiago de Chuco). Limita por el norte con parte de las Cuencas de Chao, Virú, Moche y Crisnejas; por el sur con parte de la cuenca Lacramarca, Pativilca y Fortaleza, por el este con la línea de cumbres de la Cordillera Blanca que constituye la divisoria de las aguas con la cuenca del Marañón, y por el oeste con las cuencas Nepeña, Casma, Huarney y el océano Pacífico. La cuenca del río Santa tiene una extensión de 12200 km² de la cual el 83 %, es decir 10200 km² corresponden a la cuenca húmeda, denominada así por encontrarse encima de los 2000 m.s.n.m, cota fijada como límite del área de escurrimiento superficial. La topografía es plana en la parte baja con pendientes menores al 15 %, ondulado, empinado y/o escarpado en la cordillera de los Andes con pendientes mayores del 15 %, en los valles interandinos de la parte media y alta existe áreas planas y colinosas con pendientes de 15 % a 45 %. En la parte baja tiene un valle, denominado Santa, muy importante por su contribución a la economía de la Región. Esta descripción se encuentra en [16].

4.3. Procesamiento de la información

4.3.1. Análisis de salto

Para el análisis de salto en la media a las estaciones se han agrupado en dos grupos: el

primer grupo ha sido conformado por las estaciones de Querococha, Olleros, Quillcay y Chancos y el segundo grupo ha sido conformado por las estaciones de Llanganuco, Parón, Colcas Los Cedros y Quitracsa. En el análisis visual y en el análisis de doble masa, los datos de las estaciones en estudio no presentan saltos.

4.3.2. Análisis de tendencia

La tendencia se ha evaluado comparando el coeficiente de correlación calculado (r_c) y coeficiente de correlación tabular (r_t). El análisis de tendencia se ha realizado en series libres de saltos. La estación Colcas presenta tendencias, lo cual se ha corregido.

4.3.3. Aplicación del modelamiento estocástico

El modelamiento estocástico se ha aplicado a las descargas medias anuales de las diferentes estaciones hidrográficas de la cuenca del río Santa. El modelamiento se ha realizado según la metodología indicada en la Figura 1.

4.4. Modelamiento estocástico de la serie de caudales medios anuales en la cuenca del río Santa

4.4.1. Prueba de estacionariedad

Tabla 1: Prueba de raíz unitaria de las descargas medias anuales ($\tau_c = 2.96$)

ESTACION	RIO	τ_μ
QUEROCOCHA	YANAYACO	-4.519
OLLEROS	NEGRO	-4.390
QUILLCAY	QUILLCAY	-4.704
CHANCOS	MARCARA	-3.720
LLANGANUCO	LLANGANUCO	-4.116
PARON	BUIN	-4.278
COLCAS	COLCAS	-3.530
LOS CEDROS	CEDROS	-6.084
QUITARACSA	QUITARACSA	-5.154

La estacionariedad de las series históricas de los caudales medios anuales se ha evaluado mediante la prueba de raíz unitaria Dickey–Fuller. El resultado obtenido se muestra en la Tabla 1.

Los resultados de la Tabla 1 indican que las series de descargas medias anuales en la cuenca

del río Santa son estacionarias; ya que en todos los casos el valor absoluto del t -valor estimado (τ_μ) es mayor que el valor crítico $\tau_c = 2,96$.

4.4.2. Identificación de los modelos estocásticos adecuados para los caudales promedios anuales de los afluentes del río Santa

A los datos observados que son estacionarios se han modelado observando el comportamiento de la función de autocorrelación (FAC) y la función de autocorrelación parcial.

(FACP) calculados hasta el desfase $k = 15$, lo cual se ha comparado con el comportamiento característico de estas funciones dadas por Gujarati y Porter [8]. Los posibles modelos identificados son AR(1), AR(2) y AR(3).

4.4.3. Estimación de los parámetros de los modelos identificados

El parámetro varianza del error o varianza del ruido blanco σ_ε^2 para los modelos ARMA(1,0), ARMA(0,1), ARMA(2,0), se muestra en la Tabla 2.

4.4.4. Validación del modelo estocástico

La validación o selección del modelo estocástico adecuado para cada estación hidrográfica se ha efectuado mediante las pruebas de criterio de información de Schwarz modificado (SIC). Los resultados se muestran en la Tabla 3.

Según los resultados que se observa en la Tabla 3 se define al modelo estocástico AR(1) como modelo regional adecuado para la serie de caudales medios anuales del río Santa, porque da el menor valor de SIC en 8 estaciones a excepción de la estación Parón, que da como modelo adecuado al modelo ARMA(2,0).

La Ecuación (5) es el modelo adecuado para explicar el comportamiento de las descargas medias anuales de los ríos de la cuenca del río Santa.

$$Q_t = \mu + \phi_1(Q_{t-1} - \mu) + \varepsilon_t \quad (5)$$

Los parámetros μ y ϕ_1 de la Ecuación (5) se muestran en la Tabla 4.

Tabla 2: Varianza del ruido blanco o variancia del error σ_ε^2 de las descargas medias anuales

N°	Estación	Modelo		
		ARMA(1,0)	ARMA(2,0)	ARMA(3,0)
1	QUEROCOCHA	0.1317	0.1305	0.1298
2	OLLEROS	0.7044	0.6974	0.6967
3	QUILLCAY	0.9746	0.9740	0.9119
4	CHANCOS	1.8663	1.8486	1.8366
5	LLANGANUCO	0.1326	0.1290	0.1259
6	PARON	0.2040	0.1841	0.1824
7	COLCAS	0.9699	0.9581	0.9550
8	LOS CEDROS	0.2453	0.2447	0.2424
9	QUITARACSA	2.9207	2.8032	2.6802

Tabla 3: Validación del modelo estocástico adecuado para los caudales medios anuales de los afluentes del Rio Santa

N°	ESTACION	MODELO	SIC	¿SELECCIONADO?
1	QUEROCOCHA	ARMA(1,0)	-38.392	Si
		ARMA(2,0)	-35.051	No
		ARMA(3,0)	-31.571	No
2	OLLEROS	ARMA(1,0)	30.346	Si
		ARMA(2,0)	33.649	No
		ARMA(3,0)	37.324	No
3	QUILLCAY	ARMA(1,0)	43.658	Si
		ARMA(2,0)	47.346	No
		ARMA(3,0)	48.361	No
4	CHANCOS	ARMA(1,0)	70.296	Si
		ARMA(2,0)	73.618	No
		ARMA(3,0)	77.066	No
5	LLANGANUCO	ARMA(1,0)	-38.125	Si
		ARMA(2,0)	-35.529	No
		ARMA(3,0)	-32.818	No
6	PARON	ARMA(1,0)	-20.469	No
		ARMA(2,0)	-20.958	Si
		ARMA(3,0)	-17.616	No
7	COLCAS	ARMA(1,0)	43.459	Si
		ARMA(2,0)	46.671	No
		ARMA(3,0)	50.255	No
8	LOS CEDROS	ARMA(1,0)	-12.905	Si
		ARMA(2,0)	-9.285	No
		ARMA(3,0)	-5.966	No
9	QUITARACSA	ARMA(1,0)	88.659	Si
		ARMA(2,0)	90.688	No
		ARMA(3,0)	92.563	No

Total de modelos seleccionados:

ARMA(1,0)= 8, ARMA(2,0) = 1, ARMA(3,0) = 0.

4.4.5. Evaluación del modelo seleccionado

Para verificar si el modelo seleccionado es adecuado para interpretar el comportamiento de los caudales medios anuales en los afluentes del río Santa, se han evaluado el comportamiento de

los residuos del modelo AR(1). Los residuos se calculan con la Ecuación (6).

$$\varepsilon_t = y_t - \mu - \phi_1 (y_{t-1} - \mu) \quad (6)$$

La evaluación de los residuos obtenidos con

Tabla 4: Parámetros media (μ), (φ_1) y varianza (γ_0) de las cuencas en estudio del modelo AR(1)

N°	ESTACION	μ	φ	γ_0
1	QUEROCOCHA	1.732	0.142	0.273
2	OLLEROS	4.758	0.779	0.310
3	QUILLCAY	7.090	1.026	0.224
4	CHANCOS	8.287	2.275	0.424
5	LLANGANUCO	3.031	0.149	0.328
6	PARON	2.018	0.230	0.338
7	COLCAS	5.624	1.133	0.380
8	LOS CEDROS	3.383	0.247	0.094
9	QUITARACSA	10.547	2.992	0.154

Tabla 6: Raíces de los polinomios autorregresivos del modelo AR(1)

N°	ESTACION	ϕ	$L = 1/\phi$
1	QUEROCOCHA	0.2730	3.66
2	OLLEROS	0.3100	3.23
3	QUILLCAY	0.2241	4.46
4	CHANCOS	0.4238	2.36
5	LLANGANUCO	0.3283	3.05
6	PARON	0.3382	2.96
7	COLCAS	0.3797	2.63
8	LOS CEDROS	0.0944	10.59
9	QUITARACSA	0.1543	6.48

Tabla 5: Pruebas de contraste de los residuos del modelo AR(1)

Prueba de medias, $t < t_c = 1,96$
 Análisis de correlación, $Q(M)B^2 < \chi^2(M - q) = 16,0$
 Análisis de normalidad, $g_c < g_t = 0,5811$

Estación	t	Q(M)	g_c
QUEROCOCHA	0.194	13.33	0.3768
OLLEROS	0.142	12.16	0.384
QUILLCAY	0.230	15.23	0.4352
CHANCOS	0.199	9.02	0.2126
LLANGANUCO	0.260	9.55	0.451
PARON	0.326	8.79	0.5171
COLCAS	0.197	8.56	0.3725
LOS CEDROS	0.130	8.72	0.4182
QUITARACSA	0.130	8.72	0.0659

la Ecuación (6) se muestra en la Tabla 5. Las pruebas de contraste de los residuos indicados en la Tabla 5, se observa que en todas estaciones de aforo cumplen con la condición de un proceso de ruido blanco que tiene media 0, que no hay correlación y tienen una distribución normal, por lo tanto el modelo estocástico AR(1) es adecuado para explicar el comportamiento temporal de las descargas anuales en la cuenca del río Santa.

4.4.6. Verificación de la estacionariedad del modelo seleccionado AR(1)

En el proceso de evaluación también se evalúa la estacionariedad del modelo seleccionado, mediante el análisis de las raíces de la ecuación característica que es una ecuación que está en función del operador de retardos.

Las raíces de los polinomios autorregresivos AR(1), en cada estación se muestra en la Tabla 6. La raíces L de los polinomios autorregresivos de orden 1 AR(1) mostrados en la Tabla 6, en valor absoluto son mayores que uno, por lo tanto el modelo AR(1) es estacionario en las estaciones hidrográficas estudiadas.

4.4.7. Generación de caudales medias anuales con el modelo AR(1) en los afluentes del río Santa

Después de definir el modelo adecuado al modelo AR(1), se han generado las descargas medias anuales en la cuenca del río Santa, mediante la Ecuación (5).

Tabla 7: Modelo estocástico ARMA(1, 0) en las estaciones en estudio

N°	ESTACION	Ecuación
1	QUEROCOCHA	$Q_t = 1,259 + 0,273Q_{t-1} + e_t$
2	OLLEROS	$Q_t = 3,283 + 0,310Q_{t-1} + e_t$
3	QUILLCAY	$Q_t = 5,501 + 0,224Q_{t-1} + e_t$
4	CHANCOS	$Q_t = 4,775 + 0,424Q_{t-1} + e_t$
5	LLANGANUCO	$Q_t = 2,036 + 0,328Q_{t-1} + e_t$
6	PARON	$Q_t = 1,336 + 0,338Q_{t-1} + e_t$
7	COLCAS	$Q_t = 3,489 + 0,094Q_{t-1} + e_t$
8	LOS CEDROS	$Q_t = 3,064 + 0,094Q_{t-1} + e_t$
9	QUITARACSA	$Q_t = 8,920 + 0,154Q_{t-1} + e_t$

Las ecuaciones del modelo estocástico AR(1) se muestran en la Tabla 7. Estas ecuaciones se han

Tabla 8: Estadísticas de los datos históricos y generados con AR(1) en las cuencas en estudio

Estación	Estadística	MEDIA	DES. ESTANDAR	CV	MÍN	MÁX
QUEROCOCHA	HISTORICO	1.73	0.38	0.22	0.91	2.66
	GENERADO	1.80	0.37	0.20	0.85	2.63
OLLEROS	HISTORICO	4.76	0.88	0.19	2.80	7.19
	GENERADO	4.70	1.01	-0.02	2.50	6.83
QUILLCAY	HISTORICO	7.09	1.01	0.14	4.34	8.90
	GENERADO	7.20	0.85	-0.13	5.47	8.79
CHANCOS	HISTORICO	8.29	1.51	0.18	4.23	12.28
	GENERADO	8.76	1.48	-0.15	4.27	13.02
LLANGANUCO	HISTORICO	3.03	0.39	0.13	4.27	13.02
	GENERADO	3.01	0.38	0.31	2.24	4.01
PARON	HISTORICO	2.02	0.48	0.24	0.90	3.31
	GENERADO	1.91	0.62	0.23	0.43	3.13
COLCAS	HISTORICO	5.62	1.07	0.19	3.51	7.72
	GENERADO	5.56	1.22	0.26	2.86	8.63
LOS CEDROS	HISTORICO	3.38	0.50	0.15	2.06	4.48
	GENERADO	3.27	0.62	0.52	2.15	4.99
QUITARACSA	HISTORICO	10.55	1.73	0.16	6.66	13.53
	GENERADO	10.50	1.48	0.21	7.33	13.33

obtenido reemplazando los parámetros μ y ϕ_1 de la Tabla 4 en la Ecuación (5).

4.4.8. Verificación de la bondad de ajuste del modelo estocástico AR(1)

Para la verificación de la bondad de ajuste del modelo AR (1) se comparan las estadísticas principales de los datos históricos y generados y también se han realizado la prueba de medias y la prueba de homogeneidad de variancias.

La media, desviación estándar, coeficiente de variación, valores mínimos y valores máximos de las series históricas y de las series generadas son mostradas en la Tabla 8. La prueba de medias y prueba de homogeneidad de variancias se muestra en la Tabla 9.

La media, desviación estándar, coeficiente de variación, valores mínimos y valores máximos de las series históricas y de las series generadas con el modelo AR(1) mostradas en la Tabla 8 son valores similares, es decir no hay error significativo, los cuales se confirman mediante la prueba de medias y prueba de homogeneidad de variancias mostradas en la Tabla 9, por lo tanto la series generadas con el modelo AR(1) y la series históricas provienen de la misma población en

cada estación de aforo.

Como las medias y las variancias de las descargas medias anuales de los datos históricos y generados son estadísticamente iguales, entonces se concluye que el modelo estocástico AR(1) es adecuado para explicar el comportamiento temporal de las descargas medias anuales en todas las cuencas en estudio.

4.5. Regionalización de los parámetros del modelo estocástico autorregresivo de orden 1 AR(1)

Los resultados mostrados en las Tablas 8 y 9 indican que el modelo adecuado para describir las descargas medias anuales en las subcuencas del río Santa, es el modelo AR(1). Las ecuaciones del modelo en cada estación se muestra en la Tabla 9. Para tener una ecuación estocástica regional en la cuenca del río Santa se han regionalizado los parámetros media (μ), (φ_1) y la desviación estándar del error (S_e) de los ríos en estudio.

4.6. Regionalización de las descarga media anual μ en la cuenca del río Santa

Para la regionalización de la descarga media anual se ha correlacionado las descargas medias

Tabla 9: Desviación estándar del error generados e históricos y el área de la cuenca

ESTACION	Se (OBSERVADO)	Se (ESTIMADO)
QUEROCOCHA	0.363	0.369
OLLEROS	0.839	0.809
QUILLCAY	0.987	1.105
CHANCOS	1.366	1.189
LLANGANUCO	0.364	0.453
PARON	0.452	0.285
COLCAS	0.985	1.049
LOS CEDROS	0.495	0.569
QUITARACSA	1.709	1.665

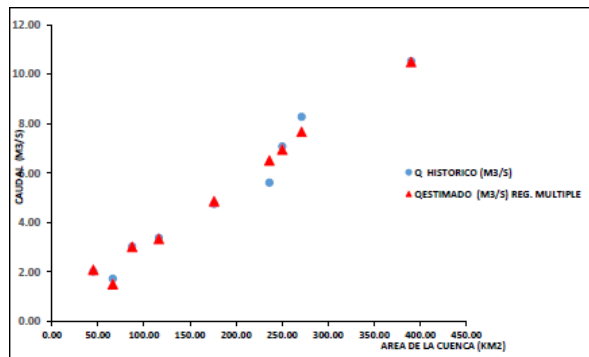


Figura 2: Correlación medias anuales, (A) área de las cuencas, (G) porcentaje de glaciares.

anuales con el área de la cuenca (A) y el porcentaje de área glaciar (G). Los resultados se muestran en la Figura 2.

En el estudio de la comparación de las medias y varianzas generadas e históricos, el área de la cuenca y porcentaje de glaciares el resultado fue que sólo en la cuenca de Colcas fue observada medias diferentes entre el histórico y el generado. Además, sólo en las cuencas de Marcana y Colcas fue observado que no hat homogeneidad entre el histórico y el generado.

La ecuación de regresión múltiple entre el caudal media anual con el área de la cuenca y el porcentaje del área glaciar es la Ecuación (7)

$$\mu_a = \bar{Q}_a = -0,3954 + 0,02745A + 0,0264G, \tag{7}$$

$$r_c = 0,9912, \quad r_t = 0,707,$$

donde:

- A = área e la cuenca en Km^2 ,
- \bar{Q}_a = descarga media anual en m^3/s ,
- $A = ?$ área de la cuenca en Km^2 ,
- G = área glaciar en porcentaje,
- r_c = coeficiente de correlación calculada,
- r_t = coeficiente de correlación tabular.

Las cuencas indicadas son los afluentes ubicadas en la Cordillera Blanca del río Santa y presenta una correlación significativa porque se tiene: $r_c > r_t$, por lo tanto el caudal medio anual (μ regional) se puede estimar mediante la ecuación (20). La correlación significativa se muestra en la Figura 2.

4.7. Regionalización del parámetro (φ_1) del modelo AR(1)

En la Tabla 4 se muestra el parámetro (φ_1) en los ríos afluentes de la cuenca del Santa, para el modelo estocástico AR(1). En la regionalización del parámetro (φ_1) no se ha encontrado correlación significativa de (φ_1) con los parámetros físicos de las cuencas, por lo tanto se ha estimado para la cuenca del río Santa como el promedio de los valores de (φ_1), cuyo valor es 0.281.

4.8. Regionalización del parámetro desviación estándar del error S_e del modelo AR(1)

La variable aleatoria ε_t del modelo regional AR(1) depende de la variancia del error σ_{ε}^2 , y esta depende del área de la cuenca.

Al correlacionar la desviación estándar del error Se con el área de la cuenca, se ha encontrado la Ecuación (8), los resultados obtenidos se muestra

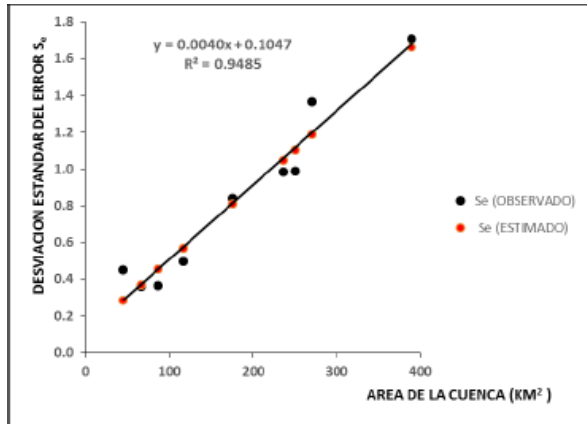


Figura 3: Correlación de la desviación estándar del error u el área de la cuenca S_e .

en la Figura 3.

$$S_e = 0,004A + 0,1047, \quad (8)$$

$r_c = 0,974$, $r_t = 0,666$, donde A = área e la cuenca en Km^2 ,

\bar{Q}_a = descarga media anual en m^3/s ,

A = área de la cuenca en Km^2 ,

G = área glaciaria en porcentaje (%),

r_c = coeficiente de correlación calculada,

r_t = coeficiente de correlación tabular.

La correlación de la desviación estándar con el área de la cuenca se obtiene resultados significativos porque $r_c > r_t$, por lo tanto desviación estándar del error (Se) regional se puede estimar mediante la Ecuación (8). La correlación significativa que existe entre el la desviación estándar Se y el área de la cuenca se observa en la Figura 3. La desviación estándar del error sirve para calcular los valores de la variable aleatoria ε_t ,

4.9. Modelo estocástico para las descargas promedio anuales de la cuenca del río Santa

El modelamiento de los caudales promedios anuales se ha ejecutado mediante el modelo estocástico AR(1). La ecuación regional de las descargas medias anuales se ha obtenido a partir de la Ecuación

$$y_t = \mu + \phi_1 (y_{t-1} - \mu) + \varepsilon_t.$$

Reemplazando las Ecuaciones regionales 7 y el valor de $(\phi_1) = 0,281$, se obtiene la Ecuación

regional (9)

$$Q_t = -0,2843 + 0,0197A + 0,01898G + 0,281Q_{t-1} + \varepsilon_t, \quad (9)$$

donde:

Q = caudal en m^3/s ,

t = tiempo en años,

A = área de la cuenca en Km^2 ,

G = porcentaje de área glaciaria ε_t = ruido blanco.

La desviación estándar del ruido depende del área de la cuenca (Ecuación (8)).

Para evaluar la prueba de bondad de ajuste del modelo AR(1) se han generado las descargas en las subcuencas en estudio mediante la Ecuación (9). La verificación de la bondad de ajuste se ha hecho mediante la prueba de medias y de variancias entre las descargas generadas y las descargas históricas.

En la verificación de la bondad de ajuste se observa que de las 9 estaciones de aforo las medias son estadísticamente iguales en 8 estaciones y en una estación las medias son estadísticamente diferentes (Colcas). En la prueba de homogeneidad de variancias en 7 estaciones son homogéneas las variables y en 2 estaciones las variables son no homogéneas (Chancos y Parón).

Los resultados obtenidos de la prueba de medias indican que las medias históricas y generadas son estadísticamente iguales en la mayoría de las estaciones, y en la prueba de homogeneidad de variancias se obtienen que las variancias de las series históricas y generadas son homogéneas en la mayoría de las estaciones. Por esta razón se concluye que la ecuación regional (22) es adecuada para estimar las descargas medias anuales en la cuenca del río Santa.

Por consiguiente en la cuenca del río Santa para generar las descargas medias anuales en cualquier punto de la cuenca, se debe conocer el área de la cuenca, porcentaje del área glaciaria, y la desviación estándar de los error que está en función del área de la cuenca (Ecuación (8)).

5. Conclusiones

Los caudales medias anuales en la cuenca del río Santa son estacionarias.

El modelo estocástico adecuado para describir el comportamiento temporal de las descargas medias anuales en la cuenca del río Santa es el modelo estocástico AR(1).

Los caudales medios anuales se pueden generar mediante el modelo regional AR(1). Los parámetros de este modelo la media μ se estima en función del área de la cuenca y del porcentaje del área glaciar. El parámetro φ_1 es independiente de los parámetros físicos de las cuencas en estudio.

Las ecuaciones regionales son aplicables para áreas de cuenca comprendidas entre 45 y 390 Km².

Las ecuaciones regionales son aplicables para áreas de glaciares expresados en porcentaje entre los límites entre 3 % y 47 %.

La desviación estándar del error S_e , para el caso del modelo estocástico AR(1) está en función del área de la cuenca.

6. Recomendaciones

Implementar programas de computación para sistematizar los cálculos de las descargas medias anuales aplicando los modelos regionales.

En el modelamiento estocástico se debe considerar la variación del área glaciar a través del tiempo.

Referencias

- [1] Jorge E Guevara Belloa, Cristóbal E Vega González, and Edilberto Guevara Pérez. Análisis estocástico de una serie cronológica de caudales mensuales. caso de estudio: Río Caroní, Venezuela. *Revista INGENIERÍA UC*, 17(1), 2010.
- [2] CL Wu and Kwok-Wing Chau. Prediction of rainfall time series using modular soft computing methods. *Engineering applications of artificial intelligence*, 26(3):997–1007, 2013.
- [3] Yan-Fang Sang. A review on the applications of wavelet transform in hydrology time series analysis. *Atmospheric research*, 122:8–15, 2013.
- [4] LM Fry, TS Hunter, MS Phanikumar, V Fortin, and AD Gronewold. Identifying streamgage networks for maximizing the effectiveness of regional water balance modeling. *Water Resources Research*, 49(5):2689–2700, 2013.
- [5] Emilio Custodio and Manuel Ramon Llamas. *Hidrología subterránea*, volume 1. Omega, Barcelona, España, 2001.
- [6] Robert S Pindyck, Daniel L Rubinfeld, and Jorge Alberto Velázquez Arellano. *Econometría: modelos y pronósticos*. McGraw–Hill, México, 4 edition, 2001.
- [7] José Hernández-Alonso. *Análisis de series temporales económicas I*. ESIC Editorial, Madrid, 2009.
- [8] Damodar-N Gujarati and DC Porter. *Econometría*. McGraw-Hill Interamericana, México, 2010.
- [9] Daniel de Rivera Peña-Sánchez. *Análisis de series temporales*. Alianza Editorial, Madrid, 2005.
- [10] Eduardo Court-Monteverde and Erick-Williams Rengifo. *Estadísticas y Econometría Financiera*. Cengage Learning, Buenos Aires, 2011.
- [11] DR. Maidment. *Handbook of applied hydrology*. McGrawHill Book Company, New York, 1993.
- [12] María Pilar González Casimiro. *Análisis de series temporales: Modelos ARIMA*, volume 04–09 of SARRIKO-ON. Economía Aplicada III/Ekonomia Aplikatua III, UPV/EHU, Bizkaia, España, 2009.
- [13] Oli GB. Sveinsson, Jose D. Salas, William L. Lane, and Donald K. Frevert. *Stochastic analysis, modeling, and simulation (SAMS) version 2007, user's manual*. Computing Hydrology Laboratory, Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, Colorado, 2007.
- [14] C Portocarrero, M Zapata, JG López, A Cochachín, G Egas, and N Santillán. Inventario de glaciares cordillera blanca. Technical report, Unidad de Glaciología y Recursos Hídricos (UGRH), Autoridad Nacional del Agua (ANA), Lima, Perú, 2010.
- [15] W. Suarez Alayza. Estudio hídrico de la cuenca del río Santa. cambio climático en la región ancash. In *Foro agua Santa*, Huaraz, Perú, 2011.
- [16] Instituto Nacional de Desarrollo. Plan de gestión de oferta de agua en la cuenca de gestión Santa. Plan regional, Ministerio de Vivienda, Construcción y Saneamiento, Lima, Perú, 2003.