Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Estudios de Postgrado
Maestría en Estadística Aplicada

MODELO ESTADÍSTICO PARA SELECCIÓN DE ESCENARIOS DE PRODUCCIÓN HIDROELÉCTRICA EN LA PLANIFICACIÓN ANUAL DEL SNI DE GUATEMALA

Ing. Byron Alberto Felipe Ajuchán
Asesorado por el Mtro. Rubén Alfredo Cerón Suchini

Guatemala, noviembre de 2021

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

MODELO ESTADÍSTICO PARA SELECCIÓN DE ESCENARIOS DE PRODUCCIÓN HIDROELÉCTRICA EN LA PLANIFICACIÓN ANUAL DEL SNI DE GUATEMALA

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

ING. BYRON ALBERTO FELIPE AJUCHÁN
ASESORADO POR EL MTRO. RUBÉN ALFREDO CERÓN SUCHINI

AL CONFERÍRSELE EL TÍTULO DE

MAESTRO EN ESTADÍSTICA APLICADA

GUATEMALA, NOVIEMBRE DE 2021

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

NÓMINA DE JUNTA DIRECTIVA

DECANA	Ing. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Armando Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez.

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANA	Ing. Aurelia Anabela Córdoba Estrada
EXAMINADOR(A)	Mtro. Ing. Edgar Darío Álvarez Cotí
EXAMINADOR(A)	Mtro. Ing. Edwin Adalberto Bracamonte Orozco
EXAMINADOR(A)	Mtro. Ing. William Eduardo Fagiani Cruz
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

MODELO ESTADÍSTICO PARA SELECCIÓN DE ESCENARIOS DE PRODUCCIÓN HIDROELÉCTRICA EN LA PLANIFICACIÓN ANUAL DEL SNI DE GUATEMALA

Tema que me fuera aprobado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 27 de enero de 2020.

Ing. Byron Alberto Felipe Ajuchán

USAC
TRICENTENARIA
Universidad de San Carlos de Guatemala

Decanato
Facultad de Ingeniería
24189101 - 24189102

DTG.725.2021

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Postgrado, al Trabajo de Graduación titulado: **MODELO ESTADÍSTICO PARA SELECCIÓN DE ESCENARIOS DE PRODUCCIÓN HIDROELÉCTRICA EN LA PLANIFICACIÓN ANUAL DEL SNI DE GUATEMALA,** presentado por el **Ingeniero Byron Alberto Felipe Ajuchán,** estudiante de la **Maestría en Estadística Aplicada,** y después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

Inga. Anabela Cordova Estrada
Decana

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
DECANO
FACULTAD DE INGENIERIA

Guatemala, noviembre de 2021.

AACE/cc

Guatemala, noviembre de 2021

LNG.EEP.OI.137.2021

En mi calidad de Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor, verificar la aprobación del Coordinador de Maestría y la aprobación del Área de Lingüística al trabajo de graduación titulado:

"MODELO ESTADÍSTICO PARA SELECCIÓN DE ESCENARIOS DE PRODUCCIÓN HIDROELÉCTRICA EN LA PLANIFICACIÓN ANUAL DEL SNI DE GUATEMALA"

presentado por **Byron Alberto Felipe Ajuchán** quien se identifica con carné **201020467** correspondiente al programa de **Maestría en artes en Estadística aplicada** ; apruebo y autorizo el mismo.

Atentamente,

"Id y Enseñad a Todos"

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA
DIRECCIÓN
ESCUELA DE POSTGRADO

Mtro. Ing. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería

Guatemala 7 de junio 2021.

M.A. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Presente

M.A. Ingeniero Álvarez Cotí:

Por este medio informo que he revisado y aprobado el Informe Final del trabajo de graduación titulado **"MODELO ESTADÍSTICO PARA SELECCIÓN DE ESCENARIOS DE PRODUCCIÓN HIDROELÉCTRICA EN LA PLANIFICACION ANUAL DEL SNI DE GUATEMALA"** del estudiante **Byron Alberto Felipe Ajuchán** quien se identifica con número de carné **201020467** del programa de Maestría en Estadística Aplicada.

Con base en la evaluación realizada hago constar que he evaluado la calidad, validez, pertinencia y coherencia de los resultados obtenidos en el trabajo presentado y según lo establecido en el *Normativo de Tesis y Trabajos de Graduación aprobado por Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014*. Por lo cual el trabajo evaluado cuenta con mi aprobación.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.

Atentamente,

MSc. Ing. Edwin Adalberto Bracamonte Orozco
Coordinador
Maestría en Estadística Aplicada
Escuela de Estudios de Postgrado

Guatemala, 20 de agosto de 2020.

Mtro. Edgar Darío Álvarez Cotí
Director de la Escuela de Estudios de Postgrado. FIUSAC.
Presente.

Estimado Maestro Álvarez Cotí:

Deseando éxitos en su labor me dirijo a su persona para manifestar lo siguiente:

Por medio de la presente hago de su conocimiento que Byron Alberto Felipe Ajuchán, estudiante de la Maestría en Estadística Aplicada, quien se identifica con carné número 2010 20467, me ha presentado el informe final de su trabajo de graduación titulado "MODELO ESTADÍSTICO PARA SELECCIÓN DE ESCENARIOS DE PRODUCCIÓN HIDROELÉCTRICA EN LA PLANIFICACIÓN ANUAL DEL SNI DE GUATEMALA", el cual realizó bajo mi asesoría, brindada en forma Ad-Honorem.

Luego de revisar el documento que contiene el informe del trabajo de investigación, manifiesto que le doy mi aprobación y considero que puede continuar con las gestiones correspondientes.

Sin otro particular, me suscribo a sus respetables órdenes.

Atentamente,

Ing. Rubén Alfredo Cerón Suchini
Maestro en Gestión Industrial

ACTO QUE DEDICO A:

Dios	Por todas sus bendiciones.
Mis padres	Reginaldo Felipe y María Yanuaria Ajuchán. Su ejemplo de esfuerzo y superación será siempre mi inspiración.
Mis hermanos	Flor de María, Rolando, Hilda, Hermelinda y Walter por su apoyo incondicional.
Mi familia	Abuelos, tíos, primos y sobrinos por su cariño y apoyo.
Farah Catú	Por su incomparable compañía y apoyo que me motiva a ser mejor cada día.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por la oportunidad brindad para continuar con formación académica.
Facultad de Ingeniería	Por permitir mi desarrollo profesional.
Ingeniero Rubén Cerón	Por su invaluable apoyo brindado para la realización de este trabajo.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

FIGURAS

TABLAS

LISTA DE SÍMBOLOS

Símbolo	Significado
GWh	Giga vatio hora
$\mathbf{m^3/s}$	Metro cúbico por segundo
$\boldsymbol{\sigma^2}$	Varianza poblacional

GLOSARIO

ARIMA	Modelo autorregresivo integrado de media móvil
Autocorrelación	Correlación de una variable consigo misma cuando se desfasa uno o más períodos de tiempo.
Bondad de ajuste	Indicado que permite discernir acerca de qué tan buena es la ecuación obtenida.
Box-plot	Nombre en inglés del gráfico o diagrama de cajas, que muestra la distancia en la que se encuentran los datos y cómo están distribuidos equitativamente.
Cuartil	Medida de posición de localización que divide una distribución de datos en cuatro partes iguales.
Distribución normal	Es la distribución de probabilidad estadística cuya función de densidad es simétrica con forma de campana.
ECMWF	Centro Europeo de Previsiones Meteorológicas a Plazo Medio.
Estocástica	Teoría estadística de los procesos cuya evolución en el tiempo es aleatoria.

Hipótesis estadística	Es una afirmación respecto a alguna característica de la población en estudio, que se formula para ser sometida a la denominada prueba de hipótesis, para ser aceptada o rechazada.
Heterocesasticidad	Varianza no constante.
Homocedasticidad	Varianza constante
LRFMME	Conjunto de modelos múltiples de pronóstico a largo plazo
Media	Se refiera a la media aritmética calculada al dividir la suma de un conjunto de datos entre el total de ellos.
Mediana	Medida de tendencia central que divide un conjunto de datos en una mitad superior y una mitad inferior.
NOAA	Administración Nacional Oceánica y Atmosférica.
Prueba de hipótesis	Técnica que permite rechazar o aceptar la hipótesis estadística en base a la información que proporciona una muestra.
Rango intercuartílico	Medida de dispersión que brinda la diferencia entre el tercer cuartil menos el primer cuartil.
Variable	Característica de una población o muestra cuya medida puede cambiar de valor.

Variable aleatoria	Es una característica considerada cuyo valor de ocurrencia sólo puede saberse con exactitud una vez observado.
Variable regresora	Variable que se usa para predecir una variable de respuesta.
Varianza	Medida de dispersión que se obtiene como el promedio de los cuadrados de las desviaciones de los valores de una variable respecto de su media aritmética.
Nivel de significancia	Probabilidad de rechazar la hipótesis nula cuando es verdadera.
SNI	Sistema Nacional Interconectado de electricidad de Guatemala.
WMO	Organización Meteorológica Mundial.

RESUMEN

El presente trabajo analiza el uso de modelos estadísticos para generar pronósticos de generación hidroeléctrica, con mayor certeza de los que actualmente se obtienen por la institución encargada de la planificación anual del sistema eléctrico de Guatemala.

El objetivo fue determinar el modelo estadístico de mejor ajuste con un nivel de confiabilidad aceptable, que permita reducir la incerteza en los pronósticos. Se analizó series de tiempo de caudales de hidroeléctricas seleccionadas para obtener modelos de tipo SARIMA, empleando la metodología Box-Jekins para pronosticar caudales, utilizados como dato de entrada a modelos de regresión que permiten determinar la energía asociada.

Con los resultados del análisis para la hidroeléctrica CHIXOY se infiere que, para centrales con embalse que permiten trasladar energía a nivel intermensual, probablemente no se tienen mejoras significativas respecto a los resultados del AMM, pero se observa un mejor seguimiento a las variaciones intermensuales de energía generada.

El análisis realizado para RENACE muestra que para centrales sin embalse o con embalse de regulación diaria, se tiene una mejora notable en los pronósticos al utilizar la metodología propuesta al reducir la incerteza del pronóstico en un 42 %.

Considerando que el 96.5 % de las centrales hidroeléctricas del país se clasifican en la segunda categoría analizada, la metodología propuesta resulta

útil para seleccionar escenarios de producción hidroeléctrica sin la necesidad de emplear software de planificación de sistemas eléctricos.

PLANTEAMIENTO DEL PROBLEMA

Contexto general

El Administrador del Mercado Mayorista realiza un programa anual de operación del sistema eléctrico de Guatemala, que incluye el pronóstico de generación de las centrales hidroeléctricas del país para lo cual utiliza modelos computacionales complejos, que permiten obtener gran cantidad de escenarios de hidrología, pero empíricamente se ha encontrado que, a pesar del uso de estos modelos, se presentan diferencias considerables entre los escenarios de planificación seleccionados y los escenarios reales de operación.

Debido a que la generación hidroeléctrica representa más del 40 % de la producción de la energía eléctrica del país, resultó de suma importancia elaborar una metodología y modelos estadísticos para pronóstico de caudales con un horizonte anual que permiten reducir la diferencia de generación hidroeléctrica entre los escenarios de planificación y operación.

Descripción del Problema

Eran requeridos modelos estadísticos confiables para la selección de escenarios de producción hidroeléctrica, basados en series históricas de caudales que permitieran reducir la incerteza de la planificación anual del Sistema Nacional Interconectado (SNI) de Guatemala.

La elaboración de los modelos implicó el análisis de las series de tiempo de los caudales y la comparación entre diferentes modelos de pronósticos, para

determinar el de mejor ajuste y utilizar estos pronósticos como variable de entrada para el modelo de estimación de generación hidroeléctrica.

Se evaluó la metodología desarrollada para algunas centrales seleccionadas, con base en su importancia debido a la energía anual generada y años de operación para tener la información necesaria para validar la certeza de los modelos. De esta manera, se estableció una metodología de aplicación general para pronósticos de generación hidroeléctrica por medio de modelos estadísticos.

Formulación del problema

Pregunta central

¿Cuál es el modelo estadístico de mejor ajuste para la selección de escenarios de generación hidroeléctrica, basado en series históricas de caudales para reducir la incerteza en la planificación anual del SNI de Guatemala?

Preguntas auxiliares

- ¿Cuál es la metodología e información que utiliza el modelo estadístico de mejor ajuste para la selección de escenarios de generación hidroeléctrica del SNI de Guatemala?

- ¿Cuál es el nivel de confiabilidad del modelo estadístico diseñado para la selección de escenarios de generación hidroeléctrica del SIN de Guatemala?

- ¿En qué medida se reduce la incerteza en la planificación anual del SNI de Guatemala con el uso del modelo estadístico diseñado?

Delimitación del problema

En el presente trabajo se desarrolló una metodología de aplicación de un modelo estadístico que permite pronosticar los caudales afluentes y la disponibilidad de generación hidroeléctrica asociada para la planificación de operación del sistema eléctrico de Guatemala en el horizonte de un año estacional entre el 1 de mayo y el 30 de abril del año siguiente para las centrales hidroeléctricas Chixoy y Renace.

El modelo de pronóstico de caudales se desarrolló con información de caudales promedios mensuales históricos de 1963 a 2018 de la base de datos de la Programación Anual de Largo Plazo 2019-2020 que realiza el AMM y para el modelo de estimación de la generación hidroeléctrica se agregó información de la generación real y pronosticada de 2006 a 2021 consignada en los informes oficiales del AMM.

OBJETIVOS

General

Elaborar el modelo estadístico de mejor ajuste para la selección del escenario de producción hidroeléctrica mediante el uso de serie de tiempo de caudales para reducir la incerteza en la planificación anual del SNI de Guatemala.

Específicos

- Establecer la metodología e información adecuada a utilizar en el modelo estadístico mediante la revisión bibliográfica y evaluación de criterios para obtener el modelo de mejor ajuste.

- Determinar el nivel de confiabilidad de predicción del modelo estadístico mediante simulaciones y comparación de escenarios para mostrar la utilidad del modelo en la selección de escenarios hidrológicos.

- Cuantificar la mejora en la predicción de la producción hidroeléctrica con el uso del modelo mediante comparación de escenarios pronosticados para mostrar el incremento en la certeza con el uso del modelo.

RESUMEN DEL MARCO METODOLÓGICO

Características del estudio

El enfoque del estudio realizado es cuantitativo ya que midió fenómenos físicos y aplicó estadística para comprender el comportamiento y analizar la realidad de forma objetiva con lo que se obtuvo un modelo de pronóstico confiable.

El alcance del estudio fue descriptivo-correlacional dado que inicialmente se describió el comportamiento y luego se analizó para correlacionar variables con las que se obtuvo un ajuste adecuado del modelo.

El diseño adoptado fue no experimental, pues la información de caudales y escenarios de generación hidroeléctrica se analizó en su estado original sin ninguna manipulación; además fue longitudinal, pues se analizó el comportamiento de los caudales y su influencia en el escenario final de producción hidroeléctrica al largo de todo el histórico disponible.

Unidades de análisis

La población que se estudió fueron los caudales afluentes a los embalses o caudales afluentes turbinados por centrales hidroeléctricas seleccionadas de Guatemala, los cuales estaban divididos en caudales promedios mensuales. Se utilizó el máximo posible u óptimo de la información histórica disponible con lo que se obtuvo la mejor información de entrada de los modelos de pronósticos.

También se estudió la energía generada asociada a los caudales de las centrales hidroeléctricas.

Variables

Las variables estudiadas se describen a continuación:

- Caudal
  - Definición teórica: cantidad de fluido que circula a través de una sección del ducto por unidad de tiempo.
  - Definición operativa: datos históricos de promedios mensuales de base de datos pública en m^3/s.
- Energía
  - Definición teórica: energía eléctrica activa generada por la central hidroeléctrica.
  - Definición operativa: energía total mensual obtenida de datos históricos reales de generación de las centrales y proyecciones de modelos en unidades de GWh .

Fases del estudio

El desarrollo del estudio se divide en 5 fases, la cuales se detallan a continuación:

Fase 1: revisión de literatura

Adicional a los antecedentes que conforman el marco referencial del estudio, se buscaron libros de texto y manuales que abordan los temas de

pronósticos con énfasis en análisis de series de tiempo y regresión lineal, incluyendo la modelización matemática, desarrollo de modelos en software y criterios de validación y selección del modelo de mejor ajuste.

Fase 2: gestión o recolección de la información

Se descargó desde la página web del Administrador del Mercado Mayorista, la base de datos de interés del software SDDP utilizada para la elaboración de la Programación de Largo Plazo del Sistema Nacional Interconectado de electricidad de Guatemala para el año 2019-2020. Esta base de datos contiene un archivo de los promedios mensuales de los caudales afluentes a todas las centrales hidroeléctricas del país con un histórico entre 15 y 69 años.

Se recolectó la información disponible de la energía eléctrica pronosticada y real de las centrales eléctricas de Guatemala para el período 2006 al 2020 desde los informes públicos oficiales del Administrador del Mercado Mayorista.

Fase 3: análisis de información

- Se seleccionó centrales hidroeléctricas de interés y se revisó la consistencia de los datos de caudales, generación pronosticada y generación real por medio de gráficas comparativas y descriptivas.

- Se utilizó la metodología Box-Jekins para la identificación del modelo de series temporales, estimar los parámetros de los modelos ARIMA de mejor ajuste para los caudales de las centrales estudiadas y comprobar las estadísticas del modelo.

- Se obtuvo modelos de regresión lineal que relacionan el caudal promedio mensual con la energía mensual generada de cada central eléctrica analizada. Se verificó los supuestos de los modelos y se seleccionaron los mejores modelos con criterios estadísticos.

- Se aplicó los modelos de regresión para obtener la generación hidroeléctrica asociada a los caudales proyectados,

- Finalmente, se realizó comparaciones entre escenarios pronosticados para mostrar la certeza del modelo planteado.

Los análisis fueron realizados utilizando el software R e InfoStat.

Fase 4: interpretación de información

Con base en los resultados obtenidos en el análisis de la información se definió el modelo de mejor ajuste para los caudales y para la generación hidroeléctrica y se obtuvieron conclusiones y recomendaciones sobre el uso del modelo de mejor ajuste.

Fase 5: redacción del informe final

En esta fase se redactó el informe final con base en los resultados obtenidos en la Fase 3 y la interpretación realizada en la Fase 4.

Técnicas de análisis de información

Para el análisis de datos se utilizó pruebas, técnicas y criterios estadísticos para describir el comportamiento de las variables, determinar los modelos de

mejor ajuste y verificar el cumplimiento de supuestos y estadísticos de los modelos.

Las técnicas de análisis de información fueron:

- Gráficos de líneas, gráficos de barras y gráficos de cajas: para describir el comportamiento de los caudales y generación hidroeléctrica.

- Gráficos de componentes de las series de tiempo: para mostrar la descomposición por el modelo aditivo, obteniendo la componente de tendencia, estacionalidad y aleatoriedad.

- Transformación de Johnson: para trasformar las series de tiempo de caudales en series con modelos SARIMA válidos.

- Test de Dickey-Fuller: para determinar la existencia de raíces unitarias de las series de tiempo y verificar las series estacionarias requeridas.

- Gráficos ACF y PACF: para evaluar las funciones de autocorrelación ACF y autocorrelación parcial PACF de las series de caudales y la autocorrelación de los residuos de los modelos ARIMA.

- Test de Shapiro-Wiks Modificado: para verificar la normalidad de los residuos de los modelos SARIMA y modelos de regresión.

- Test de Bartlett: para verificar la homogeneidad de varianza de los residuos en los modelos SARIMA y modelos de regresión.

- Test Box-Pierce: para la significancia de los rezagos en los residuos de los modelos SARIMA.

- Test de Durbin-Watson: para verificar la independencia de los residuos en los modelos de regresión.

- Histograma: para mostrar el comportamiento normal en los residuos de los modelos de regresión.

- Gráficos de dispersión: para mostrar la relación entre variables y los supuestos homocedasticidad e independencia de los residuos de los modelos de regresión.

INTRODUCCIÓN

Debido a las diferencias significativas observadas entre los pronósticos de generación hidroeléctrica, de la planificación anual realizada por el Administrador del Mercado Mayorista y los valores reales observados en los últimos años, se realizó un análisis que permite evaluar el uso de modelos estadísticos, para la selección de escenarios de producción hidroeléctrica para reducir la incerteza de los pronósticos.

En el presente estudió se determinó una metodología que utiliza modelos estadísticos, basados en series de tiempo de caudales para obtener pronósticos de generación hidroeléctrica. El análisis se realizó empleando modelos SARIMA para pronosticar caudales y modelos de regresión para obtener la energía asociada.

Los resultados obtenidos permiten conocer ventajas y limitaciones del uso de modelos estadísticos para pronosticar generación de los diferentes tipos de centrales hidroeléctricas del país.

El informe final del trabajo está compuesto por cuatro capítulos. El primer capítulo está compuesto por el marco referencial del estado del arte sobre el tema de pronóstico de caudales con enfoque en planificación de sistemas eléctricos.

El segundo capítulo presenta conceptos generales de modelos de pronósticos de tipo ARIMA y modelos de regresión, así como técnicas estadísticas para validación y selección de modelos.

En el capítulo tres se muestran los resultados de los modelos de mejor ajuste y una comparación de los pronósticos generados con la metodología actual y la metodología propuesta en este trabajo.

En el cuarto capítulo se realiza el análisis de resultados de los modelos y las comparaciones con la metodología actual.

1. MARCO REFERENCIAL

1.1. Estudios Previos

La planificación anual de los sistemas eléctricos es de suma importancia porque permite a la institución encargada de la operación del sistema garantizar la cobertura de la demanda de potencia y energía o bien prever las condiciones de déficit para la toma de las medidas necesarias.

Uno de los aspectos que provoca dificultad en la planificación es el pronóstico de los caudales a utilizar como insumos de los modelos de optimización, es por esto por lo que el pronóstico de caudales fue el enfoque de una cuidadosa revisión bibliográfica de tesis de maestría y artículos científicos interesantes para el establecimiento del estado del arte.

Con la revisión se determinó que es una rama ampliamente estudiada y abordada por medio de una gran cantidad de modelos buscan reducir el error por medio de modelos convencionales y modernos que buscan correlacionar temporal y espacialmente las características de las series temporales de caudales y variables externas que ayuden a la mejora de los pronósticos.

Antecedentes relevantes de estudios que analizan la modelación, simulación y pronósticos de hidrología que fueron útiles para el desarrollo del presente trabajo se describen en los siguientes párrafos.

El Administrador del Mercado Mayorista elabora informes anuales y semestrales sobre planificación de la operación del sistema eléctrico de

Guatemala, donde se incluye la estimación de generación hidroeléctrica para el periodo de un año y seis meses respectivamente.

En 2019, el Administrador del Mercado Mayorista desarrolló la programación de largo plazo para el período mayo 2019-abril 2020, donde evalúa las condiciones hidrológicas esperadas para el año siguiente con base en las condiciones esperadas para trimestre abril a junio de 2019, las condiciones de precipitación y realiza una estimación para el SNI.

Con base en las condiciones esperadas del primer trimestre del año de estudio analiza la evolución del fenómeno de El Niño y se establecen años análogos que posteriormente utiliza para realizar la estimación para el SNI.

Las condiciones de precipitación las analiza según el modelo NMME del Centro de Predicción Climática de la NOAA y de la WMO/LRFMME, así como el pronóstico probabilístico de ECMWF, determinando la probabilidad de precipitaciones por debajo o encima de lo normal mediante un análisis gráfico de mapas.

En el estudio se realiza la estimación teniendo en cuenta los pronósticos anteriores, sobre la tendencia por debajo o encima del promedio y utiliza un modelo estocástico que simula 50 escenarios hidrológicos mediante series sintéticas de caudales, utilizando como año inicial uno de los años análogos establecidos en la primera etapa en este caso el año 2006.

Este estudio permitió obtener los resultados de los pronósticos que se tienen con la metodología actual, así como conocer la forma, metodología y otras fuentes de información externa que se pueden tener en cuenta al elaborar un modelo.

La empresa de consultoría PSR Energy Consulting (PSR, 2018), ha desarrollado un modelo de optimización de sistemas eléctricos, enfocado en la planificación bajo incertidumbre del uso del agua en la producción hidroeléctrica, que resulta en el uso óptimo del recurso a lo largo de un período de análisis, un problema que puede ser planteado como un árbol de decisiones del uso o almacenamiento del agua y sus consecuencias futuras de acuerdo con el escenario futuro de caudales.

El modelo de optimización incluye un modelo estocástico de caudales de series sintéticas, que caracteriza la dependencia de una serie de caudales con su histórico reciente con un modelo autorregresivo periódico ARP, un vector de variables aleatorias con distribución normal calculada desde los parámetros observados de los caudales y la dependencia entre diferentes afluencias por medio de un modelo multivariado.

Adicionalmente el modelo desarrollado por PSR Energy Consulting, permite la representación de variables climáticas exógenas que permite modelar influencia de fenómenos climáticos como el niño en los resultados de los caudales.

El manual del modelo permite conocer a profundidad el enfoque de análisis estocástico y formulación matemática que emplea un software, que produce series sintéticas de caudales para estudios de planificación de mediano y largo plazo de sistemas eléctricos de potencia, lo cual fue considerado en la metodología de solución al realizar comparaciones entre diferentes enfoques de análisis que proveen otros estudios

Dmitrieva (2015) en su tesis de maestría para Ostfold Universite College, desarrolló modelos de pronósticos de la producción de energía basados en

machile learning y métodos estadísticos aplicados a series de tiempo para dos centrales hidroeléctricas de 2.4 MW y 4.0 MW respectivamente, ubicadas en Hyen, oeste de Noruega.

En este trabajo fueron construidos varios modelos como ARIMA, regresión, redes neuronales y máquinas de vectores de soporte, los cuales buscaban incluir rasgos exógenos relevantes como precipitación y temperatura. Los modelos fueron comparados en términos del error absoluto medio MAE, obteniendo como mejor método el de redes neuronales recurrentes RNN incluso mejor que el método de redes neuronales de avance.

Otro modelo de predicción con buenos resultados fue el de máquinas de vectores de soporte.

El estudio se desarrolló para pronóstico de corto plazo y determinó que el uso de variables exógenas, para un horizonte de este tipo empeora o no mejora los pronósticos obtenidos.

Este trabajo aportó información sobre los resultados que se pueden esperar entre diferentes métodos de pronóstico para corto plazo y los criterios de evaluación para seleccionar el mejor modelo. Otro aspecto importante que aportó fue la premisa que el uso de variable exógenas no necesariamente mejora el modelo, un hecho que fue tomado en cuenta al momento de considera diferentes modelos para el análisis y pronóstico de las series temporales de caudales.

Clement (2015) trabajó en un modelo regional de pronóstico de generación de centrales hidroeléctricas, a partir de pequeñas centrales de la Región NO5 en el oeste de Noruega como parte de sus tesis de maestría de la Norwegian University of Science and Technology. El estudio está enfocado en el trabajo de

planificación y operación centralizada que debe realizar el Operador del Sistema de Noruega y la importancia en tener pronósticos confiables para garantizar el abastecimiento de demanda local y la coordinación con sistemas interconectados.

El estudio se desarrolla bajo el supuesto que no es práctico crear un pronóstico individual de cada hidroeléctrica y es más útil para el Operador del Sistema crear un pronóstico agrupado para todas las hidroeléctricas de la región, enfocándose en la búsqueda de un modelo de generación combinada de todas las pequeñas centrales.

Los métodos probados fueron el de análisis regional por regresión lineal múltiple y un modelo conceptual que utiliza datos históricos individuales y obtiene datos regionales mediante la agregación de valores individuales. El mejor modelo para la predicción se decide con base en la bondad de ajuste, complejidad, requerimiento de datos y costo de operación.

El aporte de este estudio fue en la metodología de análisis y selección del modelo de pronóstico conjunto por regiones con características hídricas similares o comportamiento de generación similar en lugar de modelos individuales, un enfoque que fue utilizado para seleccionar centrales hidroeléctricas ubicadas en una misma región climática del país para comparar similitudes en el comportamiento de los caudales.

Smith, et al. (2004), desarrollaron un modelo computacional que permite apoyar el pronóstico de caudales utilizando metodologías para la proyección mensual de caudales para el sector eléctrico colombiano para etapas trimestrales, semestrales y anuales.

Las metodologías utilizadas son las Redes Neuronales Artificiales, Redes Adaptativas Neuro-Difusas, Análisis Espectral Singular, Modelo Estructural y Modelo Físico. El modelo computacional desarrollado permite agregar variables exógenas como variables macro climáticas predichas.

El uso de la herramienta computación determinó que los mejores resultados se obtienen con los modelos Estructural y Espectral, también se obtuvieron buenos resultados para algunas estaciones por medio del modelo Físico, pero este requiere buena predicción de variables exógenas como la lluvia.

Los modelos obtenidos permiten mejorar en promedio un 15 % las predicciones de los modelos triviales. El aporte de este trabajo radica en la implementación de modelos no convencionales para pronóstico de caudales mensuales con horizontes de hasta un año lo cual es consistente con la temporalidad de los modelos que se desarrollaron para el sistema eléctrico de Guatemala y brinda el antecedente de elaboración de pronósticos para el mismo horizonte y variable en estudio.

Vega (2016) realizó un análisis sobre la forma de abordar la variabilidad presente en las energías renovables como la energía eólica, solar e hidráulica, por medio de generación de series sintéticas que permiten caracterizar la función de densidad de probabilidad y la dependencia espacial y temporal de las mediciones cuando no se dispone de información suficiente.

En esta tesis de maestría se clasifican los métodos para generación de series sintéticas en modelos que consideran dependencia temporal como las Cadenas de Markov, ARMA, ARIMA, Teoría de Cópula, redes neuronales y algoritmos genéticos y modelos que consideran dependencia temporal y espacial

como Vector Autoregresivo VAR, CARMA Teoría de Cópula y Componentes principales más ARMA.

El análisis de Vega se centra en la energía solar y eólica pero las metodologías pueden ser aplicables para energía hidráulica. La metodología para selección de modelos de generación de series sintéticas propuesta incluye el análisis del tipo problema de planificación que se desea afrontar con las series, el tipo de simulación respecto a su asociación temporal, tipo de sistema y horizonte de análisis así como el tipo de recurso y fuente de información que se posee en cuanto la resolución temporal, para posteriormente realizar un análisis estadístico que incluye función de densidad de probabilidad, verificación de tendencias y estacionalidades y el análisis de dependencia temporal y espacial.

Para el análisis estadístico propone el uso de las pruebas Anderson-Darling, Kolmorov-Smirnov, Dicker-Fuller, Friedman, Kruskal Wallis y Ljung-Box aparte del análisis gráfico. Finalmente, Vega propone como trabajo futuro el desarrollo de una herramienta computacional que permita automatizar la toma de decisión debido a que uno de los principales problemas que se tiene es que los usuarios pueden ser expertos en sistemas eléctricos, pero no en modelos de generación de series sintéticas.

El trabajo de Vega aportó una metodología general sobre cómo seleccionar el mejor modelo de pronóstico en función de características temporales y espaciales de las variables a analizar, con lo cual se acotó la cantidad y tipo de modelos a analizar. Adicionalmente proporcionó una guía de las técnicas estadísticas utilizadas para el análisis de series de tiempo y de los modelos desarrollados.

En su tesis de maestría, Morales (2016) explica que, en la planificación de un sistema eléctrico con participación significativa de hidroelectricidad, la incertidumbre hidrológica es una variable que le dan al sistema un carácter estocástico. Morales expone que, al momento de elaboración del trabajo, la información hidrológica utilizada para la programación de largo plazo del sistema hidroeléctrico es de carácter histórico, utilizando años consecutivos de acuerdo con el horizonte de programación, esto supone correlaciones espaciales y temporales asociadas a los flujos y no se indaga a fondo en la autocorrelación de seris de tiempo hidrológicas.

El trabajo propone la generación de escenarios de caudales como información a utilizar en la programación de largo plazo, incorporando correlaciones espaciales y temporales y forzantes de la hidrología. La metodología tiene cierta similitud a la utilizada por Clement (2015) dado que se realiza una estimación de caudales en cuencas representativas y se utiliza un concepto de zona homogénea para poder modelar una región en la cual se realiza transposición de caudales a cuentas de hidroeléctricas interconectadas.

Dentro de los resultados relevantes obtenidos por Morales, se encuentra la importancia de toma un periodo de 30 años para representar los fenómenos en términos de caudales, ya que con esto se asegura que están presentes los fenómenos como ENSO (El Niño) y oscilaciones de carácter década como la Oscilación Decadal el Pacífico. Otro aspecto importante en el modelo obtenido se tiene con relación a la transposición de caudales, debido a que se debe verificar la correlación entre la cuenca seleccionada y los caudales observados en los puntos de interés. Dentro de los aportes importantes del trabajo de Morales al desarrollo de un modelo estadístico para el sistema eléctrico de Guatemala se tiene la premisa de cantidad de años de caudales históricos que permite

representar fenómenos estacionales y cíclicos lo cual fue utilizado para selección de estaciones hidrológicas utilizadas en el desarrollo de los modelos.

Zuñiga y Jordán (2005) realizaron un trabajo sobre pronóstico de caudales promedios mensuales usando Sistemas Neurofuzzy, donde buscaban obtener una metodología aplicada al pronóstico de caudales afluentes mensuales en centrales hidroeléctricas los cuales pueden ser utilizados como datos de entrada a los modelos de planificación de largo plazo de sistemas eléctricos.

El trabajo utiliza una combinación de Sistemas de Inferencia Fuzzy (lógica difusa) y Redes Neuronales en un modelo denominado ANFIS. El modelo aborda la cantidad y tipo de funciones de pertenencia de las variables del modelo por medio de un método heurístico.

En el trabajo se comparan los resultados del modelo y resultados de modelos de series de tiempo estocásticas para dos centrales hidroeléctricas utilizando como medida el error cuadrático medio, error porcentual medio y error absoluto medio.

Este trabajo se adaptó a los objetivos de desarrollar un modelo de pronósticos para el sistema eléctrico de Guatemala tanto en la temporalidad como en la metodología por lo que brindó un marco metodológico de comparación entre diferentes modelos incluidas algunas medidas de comparación que se utilizaron.

Moreno y Salazar (2008) presentaron un trabajo donde utilizan un modelo autorregresivo multivariado para la generar series sintéticas hidrológicas. El modelo emplea como variable de entrada las series históricas de caudales, pero tiene un enfoque de cambio al usar la temperatura superficial del mar como

variable exógena, considerando que esta variable se relaciona estadística y físicamente con los caudales. Este trabajo aportó esencialmente a la metodología de establecimiento de un modelo estadístico de pronósticos puesto que permite identificar las características de un modelo de generación de series sintéticas y evaluar cómo realizar una comparativa con otros modelos.

La modelación estocástica de caudales medios anuales en Perú fue abordada por Díaz y Guevara (2016), dando interpretación de comportamiento temporal y espacial de series estacionarias. En el trabajo se hace una completa explicación del análisis de series de tiempo de caudales y se proporciona una metodología completa desde el análisis de datos hasta la obtención y uso de modelos regionales de caudales medio mensuales.

El principal aporte de este estudio fue el establecimiento de una ruta de análisis y solución para determinar el modelo de mejor ajuste entre varios modelos, ruta que fue adaptada a los objetivos del desarrollo de un modelo de pronósticos de la generación hidroeléctrica para Guatemala, tomando en cuenta la actual metodología y su comparación con diferentes modelos generados.

Meis y Llano (2017) propusieron un modelado estadístico para el pronóstico de series de tiempo de caudales mensuales de la baja Cuenca del Plata basado en modelos estacionales autorregresivos de medias móviles SARIMA. El trabajo presenta desde análisis exploratorio de la serie de tiempo de caudales hasta el desarrollo y la selección de los mejores modelos.

En el trabajo se realizó un análisis de la serie de tiempo de 100 años por medio de subperiodos de 20 años y realiza un pronóstico para 32 meses basados en el último subperíodo analizado. Para la selección de los mejores modelos utiliza los criterios MPA, MPE, MAE, RMSE, ME y AIC.

El estudio desarrollado por Meis y Llano aportó una guía para el análisis y desarrollo de los modelos de pronósticos de caudales para las centrales hidroeléctricas de Guatemala porque desarrolla completamente la metodología Box-Jekins para la misma variable en estudio.

Uno de los aspectos relevantes al desarrollar modelos con fines de pronóstico es verificar el cumplimiento del supuesto de normalidad de los residuos porque esto garantiza que los coeficientes de los modelos sean eficientes. Razali y Wah (2011) compararon la potencia de las pruebas de normalidad Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors y prueba de Anderson-Darling concluyendo que la prueba más poderosa es la de Shapiro-Wilk.

En el desarrollo del trabajo se expone que esta prueba inicialmente estaba restringida para tamaños de muestra menores a 50, pero se han realizado modificaciones que permiten su uso para muestras de tamaño 5000. Este trabajo es de suma importancia para el trabajo porque indica qué prueba utilizar al momento de realizar el diagnóstico de los modelos de pronóstico.

Como se ha descrito anteriormente, existe un abordaje interesante y variado al problema de predicción de caudales y energías renovables variables en la planificación de sistemas eléctricos.

Existe una amplia gama de modelos que han sido probados y analizados, así como criterios y metodologías que buscan obtener la mejor caracterización del comportamiento de los caudales en diferentes enfoques, periodos de planificación y complejidad los cuales pueden ser utilizados como base para nuevos estudios.

Una de las principales tendencias metodológicas es el tratamiento estocástico del problema y solución por medio de series sintéticas, una metodología muy extendida para procesos de planificación de mediano y largo plazo (de uno o más años) sin embargo, también se observa que se mantiene el uso de modelos autorregresivos que analizan series temporales históricas de los caudales para pronósticos hasta de un año.

Para el corto plazo se mantiene el uso de modelos autorregresivos y nuevos modelos basados en machine learning. Los antecedentes muestran un interés relevante de inclusión de variables exógenas como temperatura y precipitación con pronósticos independientes en busca de mejorar los resultados.

1.2. Marco contextual

En Guatemala existe un marco legal y normativo del funcionamiento del mercado y sistema eléctrico conformado por la Ley General de Electricidad y su Reglamento, El Reglamento del Administrador del Mercado Mayorista y las Normas de Coordinación Comercial y Operativa del Administrador del Mercado Mayorista.

Esta regulación faculta al Administrador del Mercado Mayorista para coordinar la operación del sistema de generación del país dentro del cual se encuentran las principales hidroeléctricas que abastecen de potencia y energía al país y permiten exportaciones de energía a México y Centroamérica. Este trabajo se desarrolló alrededor de la información generada por el AMM en cuanto a pronósticos y operación real de las centrales hidroeléctricas del país para un periodo de un año.

2. MARCO TEÓRICO

2.1. Fundamentos estadísticos de modelos de pronósticos

Los métodos de pronósticos clásicos se pueden dividir en cualitativos, de proyección histórica y causales. Estos métodos se diferencian en el fundamento lógico de uso de información y experiencia, la complejidad del análisis cuantitativo y la exactitud de los pronósticos para corto y largo plazo (Ballou, 2004).

Un método cualitativo se caracteriza por estimar valores cuantitativos futuros usando el juicio, técnicas comparativas, la percepción o encuestas. Tienen una naturaleza no científica que dificulta su estandarización y la validación de su precisión (Ballou, 2004).

Los métodos de proyección histórica consideran que los valores futuros se pueden replicar en función de valores pasados al menos en alguna parte. El pronóstico se realiza adecuadamente si se tienen series temporales definidas.

El uso de modelos estadísticos y matemáticos como herramienta en los pronósticos de series de tiempo se fundamenta en las características cuantitativas de éstas. Estos modelos rastrean los cambios al ser actualizados a medida que se dispone de nueva información, característica que le permite adaptarse a los cambios en patrones de tendencia y estacionales (Ballou, 2004). Estos métodos tienen buenos resultados en periodos de corto y mediano plazo.

Entre los modelos utilizados se encuentran: promedios móviles y suavizamiento exponencial los cuales tienen el inconveniente de no capturar de forma adecuada toda la influencia presente en los datos históricos, así como modelos ARIMA que incluyen características de una serie de tiempo usando su autocorrelación.

La construcción de métodos causales considera que los valores de una variable que se busca pronosticar se derivan de los valores de otras variables asociadas. En la medida que se puedan definir adecuadas relaciones causa efecto, un modelo causal permite anticipar mayores cambios en una serie de tiempo y obtener pronósticos precisos para un periodo de mediano y largo plazo.

La dificultad de encontrar variables causales reales es uno de los inconvenientes en el desarrollo de este tipo de modelos (Ballou, 2004). Ejemplo de estos modelos son los de regresión lineal los cuales caracterizan tendencias lineales que permiten usar una o más variables explicativas para realizar los pronósticos.

Este trabajo se centró en el análisis de series de tiempo de caudales históricos promedios mensuales, para obtener el modelo de pronóstico de tipo ARIMA de mejor ajuste y un modelo de estimación de energía asociada por medio de modelos de regresión, por lo que la teoría desarrollada a continuación se centra en estos métodos.

2.1.1. Análisis de series de tiempo

El proceso de desarrollar un pronóstico inicia con la recolección de datos anteriores ocurridos durante varios períodos. El conjunto de datos resultantes se denomina serie de tiempo o serie temporal. Los períodos de tiempo pueden ser

anuales, trimestrales, mensuales, diarios u horarios. Uno de los principales usos del análisis de series de tiempo es la predicción de valores futuros de una variable que forma parte de la serie de tiempo a partir de observaciones pasadas (Webster, 2001).

2.1.1.1. Series de tiempo y sus componentes

La mayoría de los métodos de pronóstico estadístico se basa en una serie de observaciones secuenciales en el tiempo o en un intervalo fijado de una variable de interés (Hillier y Lieberman, 2010).

Si una variable X_i es la variable aleatoria de interés en el tiempo i y si las observaciones se toman en los momentos $i = 1, 2, \dots, t,$ entonces los valores observados $\{X_1 = x_1, X_2 = x_2, \dots, X_t = x_t,\}$ son una serie de tiempo. (Hillier y Lieberman, 2010).

La mayoría de las series son muy complejas y todas contienen por lo menos uno de los siguientes cuatro componentes: tendencia secular, variación estaciona, variación cíclica y variación irregular o aleatoria (Webster, 2001):

- Tendencia secular: es la conductora en el largo plazo de la variable para un período de longitud prolongada. Refleja la dirección predominante como ascendente o descendente.
- Componente estacional: las fluctuaciones estacionales que componen una serie de tiempo son patrones que tienden a ocurrir de nuevo, regularmente, durante algún período.
- Variaciones cíclicas: corresponden a los cambios por arriba y por debajo de la tendencia en un período prolongado. Abarcan períodos mucho más

prolongados que las variaciones estaciones y con frecuencia incluyen tres o más años de duración.

- Variaciones irregulares: son variaciones aleatorias originadas por sucesos atípicos que generan movimientos con un patrón no identificable.

2.1.1.2. Modelos de series de tiempo

Un modelo de serie de tiempo puede expresarse como alguna combinación de los cuatro componentes vistos anteriormente. El modelo es simplemente una expresión matemática de la relación entre los cuatro componentes. Comúnmente dos tipos de modelos se relacionan con las series de tiempo: modelo aditivo y modelo multiplicativo (Webster, 2001):

Webster (2001) expresa el modelo aditivo como:

$$Y_t = T_t + S_t + C_t + I_t \qquad \text{(Ec. 1)}$$

Donde:

Y_t es el valor de la serie temporal en el periodo t.

T_t es la tendencia.

S_t es la variación estacional.

C_t es la variación cíclica.

I_t es la variación aleatoria.

Webster (2001) indica que el modelo aditivo conlleva el supuesto que los cuatro componentes de la serie son independientes unos de otros, cuando en realidad muchas veces el cambio en una de las componentes genera cambios en otros componentes incumpliendo con el supuesto de independencia. Por este motivo se prefiere el uso del modelo multiplicativo frecuentemente, en el cual los

componentes interactúan entre sí de forma dependiente. El modelo multiplicativo se expresa como (Webster, 2001):

$$Y_t = T_t \times S_t \times C_t \times I_t \qquad \text{(Ec. 2)}$$

El modelo multiplicativo se expresa T_t en unidades originales y las componentes S_t, C_t e I_t se expresan en valores de porcentajes. En series de datos anuales suele desaparecer la componente S_t (Webster, 2001).

2.1.1.3. Descomposición de la serie de tiempo

En Webster (2001), se indica que muchas veces es útil descomponer una serie de tiempo en sus cuatro componentes para evaluar cada componente de forma individual, dado que cada componente puede brindar información relevante y útil para el modelado. El análisis individual de las componentes realizado por Webster (2001) se presenta a continuación:

- Aislamiento de la componente estacional: consiste en desarrollar un índice estacional por medio de un promedio móvil centrado. La razón del promedio móvil se obtiene dividiendo el valor original Y_t por el promedio móvil PM. El resultado produce los componentes S_t e $I_t$de la serie de tiempo como lo muestra la siguiente ecuación:

$$\frac{Y_t}{PM} = \frac{T_t \times S_t \times C_t \times I_t}{T_t \times C_t} = S_t \times I_t \qquad \text{(Ec. 3)}$$

Posteriormente, se determina la razón media por promedio móvil para cada periodo. Estas razones se deben normalizar dividiendo el número de periodos por la suma de las razones promedio por promedio móvil, obtenido la denominada razón de normalización. Al multiplicar la razón de

normalización por la razón promedio se obtiene el índice estacional para cada periodo. Estos índices se pueden utilizar para desestacionalizar los datos eliminando las variaciones estacionales. Estos valores sin estacionalidad denominados valores corregidos estacionalmente se determinan dividiendo el valor real por el índice estacional.

- Aislamiento de la variación cíclica: las componentes cíclicas e irregulares se obtienen al dividir los datos originales por la norma estadística la cual contiene T_t y S_t como indica la ecuación 4, dado que $Y_t = T_t \times S_t \times C_t \times I_t$:

$$\frac{Y_t}{T_t \times S_t} = \frac{T_t \times S_t \times C_t \times I_t}{T_t \times S_t} = C_t \times I_t \qquad \text{(Ec. 4)}$$

- Aislamiento de la variación irregular: con frecuencia es posible suavizar o eliminar estas componentes de manera efectiva utilizando un promedio móvil.

2.1.2. Pronósticos con series de tiempo

Gujarati y Porter (2010) explican que uno de los enfoques de pronósticos con series de tiempo, son los modelos autorregresivos integrados de medias móviles (ARIMA) desarrollado por medio de la metodología conocida como Box-Jekins. El enfoque de estos modelos es que una variable dependiente, no se explica a través de variables regresoras diferentes sino se explica por valores rezagados o anteriores de ella misma y por una componente estocástica de error.

Dada la naturaleza del trabajo, la descripción se hace para modelos ARIMA que pertenecen a una sola serie de tiempo conocidos como ARIMA univariados.

2.1.2.1. Procesos autorregresivos (AR), procesos de medias móviles (MA) y ARMA

Gujarati y Porter (2010), indican que si una variable tiene valor Y_t en un periodo t, se puede modelar el valor de Y en el tiempo t como un proceso estocástico autorregresivo de orden p o AR (p) donde su valor depende de valores de periodos pasados como una proporción α de su valor desde el periodo $(t-1)$ hasta el periodo $(t-p)$ más una componente aleatoria u_t con los valores de Y expresados alrededor de su media δ de la siguiente forma:

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \cdots + \alpha_p(Y_{t-p} - \delta) + u_t \qquad \text{(Ec. 5)}$$

Otra forma de generar el valor de de Y en el tiempo t, es como una combinación lineal de valores de error de ruido blanco, donde se obtiene el valor de Y igual una constante μ más un promedio móvil del error u presente y pasado desde el periodo t hasta el periodo $(t-q)$ siendo q el orden del proceso MA (q). El valor de Y_t se escribe como (Gujarati y Porter, 2010):

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \cdots + \beta_q u_{t-q} \qquad \text{(Ec. 6)}$$

Si el valor de Y tiene características de proceso autorregresivo (AR) y (MA) al mismo tiempo se considera con un proceso ARMA que tendrá p valores autorregresivos y q valores de promedio móvil y será abreviado como $ARMA$ (p, q). La expresión de este proceso está dada por:

$$Y_t = C + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} + \cdots + \alpha_p Y_{t-p} + \beta_q u_{t-q} \qquad \text{(Ec. 7)}$$

Donde:

Y_t es la última observación disponible.

u_t error del ruido blanco.

p es el orden autorregresivo de la serie.

q es el orden de la componente de media móvil.

α_p coeficiente autorregresivo de la serie.

β_q coeficiente de media móvil.

C término constante.

2.1.2.2. Método de Box-Jekins

En la práctica frecuentemente se eligen métodos de pronóstico, sin verificar adecuadamente si el modelo es apropiado para la aplicación. El método Box Jekins coordina cuidadosamente el modelo y el procedimiento. Este método emplea un enfoque sistemático para identificar un modelo, elegido de una rica clase de modelos y genera un procedimiento de pronóstico apropiado. El método es de naturaleza iterativa que requiere una gran cantidad de datos pasados por lo menos 50 periodos de tiempo. En el corto plazo este método resulta ser mejor que la mayoría de los otros métodos de pronósticos (Hillier y Lieberman, 2010).

El procedimiento del método Box Jekins es explicado por Hillier y Lieberman (2010) de la siguiente manera:

- Primero se elige un modelo por medio del cálculo de autocorrelaciones y autocorrelaciones parciales y evaluación de los patrones. Una autocorrelación mide la correlación entre valores de series de tiempo separado por un número fijo de periodos denominado retraso, por lo tanto, la autocorrelación para un retraso de dos periodos mide la correlación entre cada otra observación, es decir la correlación entre la serie de tiempo original y misma serie de tiempo que avanzó dos períodos. Una autocorrelación parcial es una autocorrelación condicional entre una serie

original y la misma serie que avanzó un número fijo de períodos, manteniendo el efecto de los otros tiempos rezagados fijos. Con el uso de software se pueden calcular las autocorrelaciones que resultan ser buenas a medida que se tiene mayor cantidad de datos.

- A continuación, se estiman los parámetros asociados al modelo utilizando el histórico de datos.

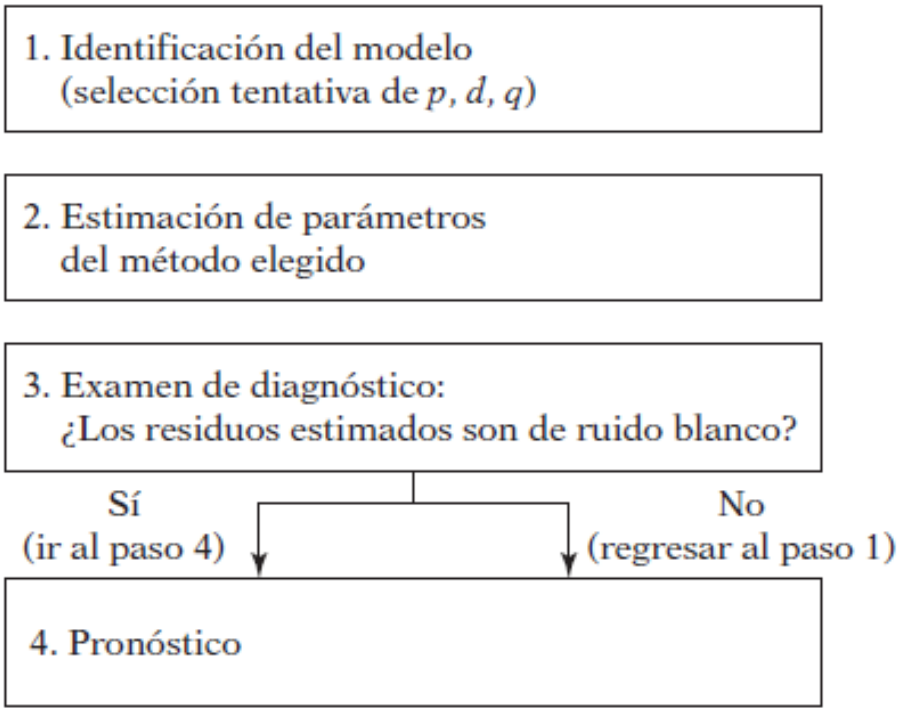
- Finalmente se pueden calcular los residuos o errores del pronóstico, cuando el pronóstico se hace retrospectivamente con los datos históricos para examinar su comportamiento. De igual manera se puede evaluar el comportamiento de los parámetros. Si el comportamiento de residuos y parámetros estimados tienen un comportamiento esperado el modelo será validado.

- El procedimiento debe repetirse hasta que se logre la validación del modelo.

Hillier y Lieberman (2010) indican que en general los residuos y los parámetros deben comportarse de manera predecible. Los residuos de la muestra deben comportarse aproximadamente como variables aleatorias, distribuidas normalmente independientes cada una con una media 0 y varianza σ^2. Por su parte, los parámetros estimados no deben estar correlacionados y ser significativamente diferentes.

La metodología de Box-Jekins descrita en los puntos anteriores se muestra en la siguiente figura:

Figura 1. **Metodología de Box-Jekins**



Fuente: Gujarati y Porter (2010). *Econometría* (p. 778).

2.1.2.3. Series estacionarias

En la etapa de identificación del modelo de la metodología Box-Jekins, debe evaluarse la estacionariedad de la serie para que sea adecuado el modelo ARIMA generado. Gujarati y Porter (2010) indican que una serie de tiempo es estacionaria si su media, varianza y auto covarianza para todos los rezagos permanecen invariantes en el tiempo.

Una de las pruebas estadísticas más utilizada sobre la estacionariedad de una serie se conoce como prueba de raíz unitaria. La existencia de una raíz unitaria significa que el coeficiente que explica la una variable en función de su

variable rezagada en $t-1$ es igual a uno en este caso se dice que la serie es o estacionaria.

La prueba de Dickey Fuller aumentada, permite evaluar las hipótesis H_0: $\delta = 0$ (la serie de tiempo es no estacionaria por existencia de raíz unitaria) y $H_1: \delta < 0$ (la serie de tiempo es estacionaria por no existir raíz unitaria) por lo cual se busca rechazar H_0 para desarrollar un modelo ARIMA.

2.1.2.4. Criterios de selección de modelos

Luego de conocer diferentes métodos de pronósticos basados en el análisis de series temporales es necesario conocer métricas o indicadores para la selección método apropiado para cada aplicación.

Identificar el modelo de mejor ajuste para la serie de tiempo y evaluar la estabilidad de los parámetros estimados son pasos importantes para la selección de un método. Sin embargo, la elección final entre varios métodos necesita de alguna medida de rendimiento. Dado que el objetivo es generar pronósticos que sean los más precisos posibles, es natural basar las medidas de rendimiento en los errores del pronóstico o también llamados residuales (Hillier y Lieberman, 2010).

El error de pronóstico es el valor absoluto de la diferencia del pronóstico respecto del valor observado real de la serie de tiempo. Por tanto, el error se determina como (Hillier y Lieberman, 2010):

$$E_t = |x_t - F_t| \qquad \text{(Ec. 8)}$$

Donde:

E_t es el error de pronóstico en el período t.

x_t es el valor real de la variable.
F_t es el pronóstico generado con el método utilizado.

Dos medidas utilizadas para el rendimiento son la desviación media absoluta (MAD) y el error cuadrado medio (MSE).

El MAD o MAE es simplemente el promedio de los errores, para evaluar n periodos de pronóstico se determina como (Hillier y Lieberman, 2010):

$$MAD = \frac{\sum_{t=1}^{n} E_t}{n} \qquad \text{(Ec. 9)}$$

El MSE o RMSE en cambio promedia el cuadrado de los errores de pronóstico como (Hillier y Lieberman, 2010):

$$MSE = \frac{\sum_{t=1}^{n} {E_t}^2}{n} \qquad \text{(Ec. 10)}$$

Hillier y Lieberman (2010) explican que la ventaja del MAD es la simplicidad del cálculo, mientras que la ventaja del MSE es que impone una penalización grande para un gran error de pronóstico que puede tener consecuencias importantes en las decisiones tomadas con base en el modelo y que en la práctica los estadísticos prefieren el MSE mientras que las personas con menor conocimiento estadístico basan sus decisiones en el MAD.

En la práctica normalmente se calculan periódicamente las medidas de rendimiento, para controlar que tan bien está funcionado el método de pronóstico y definir la necesidad de nuevos métodos.

Otro criterio utilizado para evaluar el ajuste de los modelos es el Porcentaje de Error Medio Absoluto MAPE, el cual mide la exactitud del pronóstico comparado con la serie de tiempo y expresa el error en valores de porcentaje. Este indicador se calcula como (Cárdenas et al., 2014):

$$MAPE = \frac{1}{t}\sum_{t=1}^{t}\left|\frac{F_t - x_t}{x_t}\right| \qquad \text{(Ec. 11)}$$

Donde:

t es la cantidad de observaciones.

x_t es el valor real observado de la variable.

F_t es el pronóstico generado con el método utilizado.

Finalmente se tiene otros criterios utilizados para selección de modelos como el Criterio de Información de Akaike AIC y el Criterio de Información de Schwarz (BIC) los cuales incorporan un factor de penalización al adicionar variables regresoras al modelo. Estos criterios se definen como (Gujarati y Porter, 2010):

$$AIC = e^{2k/n}\frac{\sum_{t=1}^{n} {E_t}^2}{n} \qquad \text{(Ec. 12)}$$

$$BIC = n^{k/n}\frac{\sum_{t=1}^{n} {E_t}^2}{n} \qquad \text{(Ec. 13)}$$

Donde:

k es el número de regresoras.

n es el número de observaciones.

Gujarati y Porter (2010) explican que el criterio BIC impone una penalización más alta que la impuesta por el AIC. Para ambos criterios se prefiere el modelo

que tenga el valor menor y ambos criterios permiten comparar el desempeño de los pronósticos de los modelos dentro y fuera de la muestra.

2.1.2.5. Examen diagnóstico de modelos ARIMA

Para verificar el comportamiento de los residuos del modelo como una variable aleatoria se pueden utilizar las siguientes pruebas estadísticas:

a) Prueba de Shapiro Wilks modificada: esta prueba permite evaluar las hipótesis H_0: los datos tienen una distribución normal y H_1: los datos no tienen una distribución normal. La prueba de Shapiro-Wilks inicialmente fue restringida para muestras con $n < 50$ pero con las modificaciones realizadas por Mahibbur y Govindarajulu (1997) se puede usar para $n \leq 5000$. El estadístico de prueba de Shapiro-Wilks W para una muestra aleatoria ordenada $x_1 < x_1 < x_n$ se define por medio de (Razali y Wah, 2011):

$$W = \frac{\sum_{i=1}^{n}(a_i y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad \text{(Ec. 14)}$$

$$a_i = (a_i, \dots a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}} \qquad \text{(Ec. 15)}$$

Los valores $m = (m_1, \dots, m_n)^T$ son los valores medios del estadístico y V es la matriz de covarianza de los estadísticos. Los valores pequeños suponen el rechazo de la normalidad y lo valores cercanos a uno la normalidad en los datos.

b) Prueba de Bartlett: se utiliza para comprobar la homogeneidad de la varianza. Las pruebas de hipótesis correspondientes son $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$ y $H_1: \sigma_1^2 \neq \sigma_2^2 \neq \sigma^2$ con las cuales se busca probar que cada varianza muestral de un conjunto $s_1^2 = s_2^2 = s_k^2$ de k muestras independientes provienen de una población normal con varianza σ^2. La hipótesis se prueba con la razón A/B con distribución aproximada con $f = k - 1$ grados de libertad donde A y B se determinan por medio de (Gujarati y Porter, 2010):

$$A = \sum_{i=1}^{k} f_i \text{ x } \ln \frac{\sum f_i s_i^2}{\sum_{i=1}^{k} f_i} - \sum (f_i \ln s_i^2) \qquad \text{(Ec. 16)}$$

$$B = 1 + \frac{1}{3(k-1)} \left[\sum \left(\frac{1}{f_i} \right) - \frac{1}{\sum_{i=1}^{k} f_i} \right] \qquad \text{(Ec. 17)}$$

c) Prueba de Box y Pierce: permite para probar la hipótesis que el conjunto de todos los coeficientes de autocorrelación muestrales ρ_k hasta el rezago m son simultáneamente iguales a 0. Las hipótesis correspondientes a esta prueba son: H_0: $\rho_1 = \rho_2 = \rho_k = 0$ y H_1: Algún $\rho_k \neq 0$. Para probar la hipótesis se utiliza el estadístico Q que se define como (Gujarati y Porter, 2010):

$$Q = n \sum_{k=1}^{m} \rho_k^2 \qquad \text{(Ec. 18)}$$

Donde:

n es el tamaño de la muestra

m es la longitud del rezago

ρ_k es la razón entre la covarianza muestral en k y la varianza muestral.

2.1.2.6. Transformación de Johnson

Lagos y Vargas (2003) explican que, cuando no se cumple el supuesto de normalidad se pueden aplicar transformaciones a los datos para convertirlos en datos normales a través del sistema de Familias de distribuciones de Johnson. En este trabajo se utilizó la familia S_B que se refiere a la variable acotada. Para una variable x la transformación aplicada a los datos pertenecientes a la familia S_B con parámetros γ, η, λ y ε es:

$$Z_B = \gamma + \eta \ln\left(\frac{x-\varepsilon}{\lambda+\varepsilon-x}\right) \qquad \text{(Ec. 19)}$$

sujeta a:
Condiciones de parámetros: $\eta > 0, -\infty < \gamma < \infty, -\infty < \varepsilon < \infty$
Condiciones de variable: $x > \varepsilon$

2.1.3. Pronósticos causales con regresión lineal

Los modelos de regresión se generan a partir de la asociación entre variables cuantitativas o cualitativas, teniendo una o más variables independientes que permite explicar el valor de una variable dependiente. Dos de los principales objetivos de los modelos de regresión son (Walpole, Myers, Myers y Ye, 2012):

- Determinar efectos de las variables independientes sobre las variables dependientes.
- Describir y pronosticar a partir de valores de una variable independiente los valores de una variable dependiente.

2.1.3.1. Modelo de regresión

El modelo de regresión lineal simple establece la asociación de una variable respuesta Y y una variable regresora o explicativa x. Una forma razonable de la relación entre Y y x es la relación lineal simple (Walpole et al., 2012):

$$Y = \beta_0 + \beta_1 x \qquad \text{(Ec. 20)}$$

Donde:
β_0 es la intersección.
β_1 es la pendiente.

Es posible tener una relación exacta determinista sin componente probabilístico entre dos variables, sin embargo, en diversos fenómenos la relación no tiene esta característica, por tanto, una variable x dada no produce el mismo valor de Y en todas las ocasiones. El análisis de regresión busca encontrar la mejor relación entre Y y x identificando la magnitud de la relación por medio de métodos que permiten obtener valores de la variable independiente dado los valores del regresor x.

Un parte importante al realizar el análisis de regresión es estimar los coeficientes de regresión β_0 y β_1. Si se denota los estimadores b_0 para β_0 y b_1 para β_1 entonces, la recta de regresión ajustada, o estimada, es dada por (Walpole et al., 2012):

$$\hat{y} = b_0 + b_1 x \qquad \text{(Ec. 21)}$$

Donde:
$\hat{y}$ es el valor pronosticado o ajustado.

b_0 estimador de la intersección β_0.
b_1 estimador de la pendiente β_1.

Esta recta ajustada es en realidad un estimado de la recta de regresión verdadera. Cuando se dispone de una mayor cantidad de datos es posible lograr que la recta ajustada se encuentra más cerca de la recta de regresión real.

Cuando se emplea más de una variable independiente que permite explicar Y se denomina regresión lineal múltiple. Para el caso de k variables independientes $x_1, x_2, \dots, x_k$ el modelo que da la media de $\mu_Y|x_1, x_2, \dots, x_k$ tiene la siguiente estructura (Walpole et al., 2012):

$$\mu_Y|x_1, x_2, \dots, x_k = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \qquad \text{(Ec. 22)}$$

Y la respuesta se obtiene de la ecuación de regresión muestral (Walpole et al.,2012):

$$\hat{y} = b_0 + b_1 x_1 + \cdots + b_k x_k \qquad \text{(Ec. 23)}$$

Donde:
$\hat{y}$ es el valor pronosticado o ajustado.
b_0 estimador de β_0.
b_1 estimador de β_1.
b_k estimador de β_k.

2.1.3.2. Regresión lineal por segmentos

Gujarati y Porter (2010) presentan el uso de variables dicótomas para obtener una regresión lineal por segmentos que consta de dos partes lineales

obtenidas por el cambio de estructura del comportamiento de la pendiente entre dos variables a partir de un valor umbral X^* de la variable regresora. La recta de regresión estará determinada por:

$$Y_i = \alpha_1 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + e_i \qquad \text{(Ec. 24)}$$

Donde:

Y_i es la variable regresada.

X_i es la variable regresora.

X^* es el valor umbral de la variable regresora para el cambio de estructura.

D_i es variable dicótoma con valor 1 si $X_i > X^*$ y valor 0 si $X_i < X^*$.

e_i es el error aleatorio.

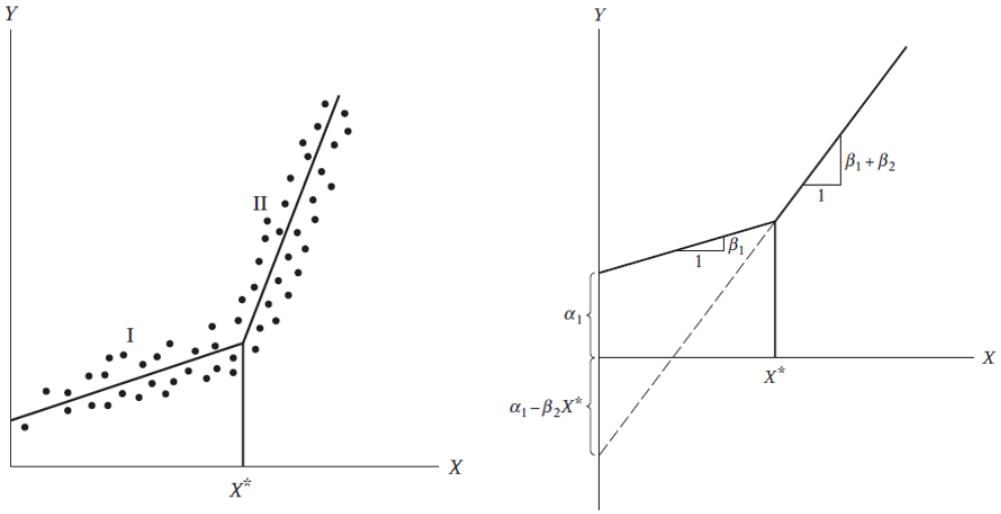
Si $E(e_i) = 0$ las rectas para cada segmento estarán definidas como (Gujarati y Porter, 2010):

$$E(Y_i | D_i = 0, X_i, X^*) = \alpha_1 + \beta_1 X_i \qquad \text{(Ec. 25)}$$

$$E(Y_i | D_i = 1, X_i, X^*) = \alpha_1 - \beta_2 X^* + (\beta_1 + \beta_1) X_i \qquad \text{(Ec. 26)}$$

El modelo expuesto anteriormente puede observarse en la figura 2 que permite visualizar la relación entre las pendientes de cada segmento, así como la conformación de los interceptos de las rectas de regresión:

Figura 2. **Relación y parámetros en regresión lineal por segmentos**



Fuente: Gujarati y Porter (2010). *Econometría* (p. 296).

2.1.3.3. Bondad de ajuste de la recta de regresión

Aparicio, Martínez y Morales (2004) plantean el uso de un estimador de la varianza σ^2 como una medida de la bondad de ajuste de un modelo de regresión lineal puesto que σ^2 es una medida de la heterogeneidad entre los individuos respecto a la media. Es posible obtener el estimado de la varianza con base en la suma de cuadrados de los residuos SSE. Para un modelo de regresión con dos parámetros se tienen $n - 2$ grados de libertad y se puede definirl el cuadrado medio residual MSE como un estimador de σ^2 como:

$$s^2 = MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} \qquad \text{(Ec. 27)}$$

Donde:

y_i son las observaciones de la variable regresada.

$\hat{y}_i$ son las estimaciones que proporciona la recta de regresión.

n es el número de observaciones de la muestra.

Aparicio et al. (2004) también presentan el coeficiente de determinación R^2 como un estadístico para evaluar la bondad de ajuste de la recta de regresión. Este coeficiente está definido como la proporción de la varianza explicada por la recta de la regresión como:

$$R^2 = \frac{SSR}{SCT} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i)^2} \qquad \text{(Ec. 28)}$$

El valor del coeficiente de determinación estará entre $0 \leq R^2 \leq 1$ implicando que para valores cercanos a 1 (entre 0.6 y 1) la varianza está explicada en Buena medida por la recta de regresión. Valores de R^2 cercanos 0 implican que la recta de regression no permite explicar la varianza. Al comparar modelos se prefiere el modelo con R^2 más alto.

2.1.3.4. Diagnóstico del modelo de regresión

Aparicio et al. (2004) explican que posterior a realizar el ajuste del modelo de regresión y superar las pruebas de bondad de ajuste se debe verificar que el modelo satisface las hipótesis o supuestos básicos del modelo los cuales son la linealidad entre las variables regresoras y regresadas, así como los siguientes supuestos de los residuos:

- Normalidad: los residuos deben tener una distribución normal con media 0 y varianza constante que representa una variable aleatoria.

- Varianza constante: para verificar que las observaciones fueron obtenidas de una misma población y su variabilidad respecto de sus medias está dada por σ^2.
- Independencia de los residuos: para probar que los valores de las observaciones de la variable respuesta no afectan unas a otras.

Como lo presenta Aparicio et al. (2004) los residuos se pueden analizar por métodos gráficos y por pruebas estadísticas. Algunas gráficas utilizadas para evaluar los residuos son:

- Histograma de residuos para evaluar el ajuste a la distribución normal con media 0.
- Gráfico de residuos versus valores de la variable regresada para descartar heterocedasticidad.
- Gráfico secuencial de los residuos (residuos versus casos) para evaluar la incorrelación.

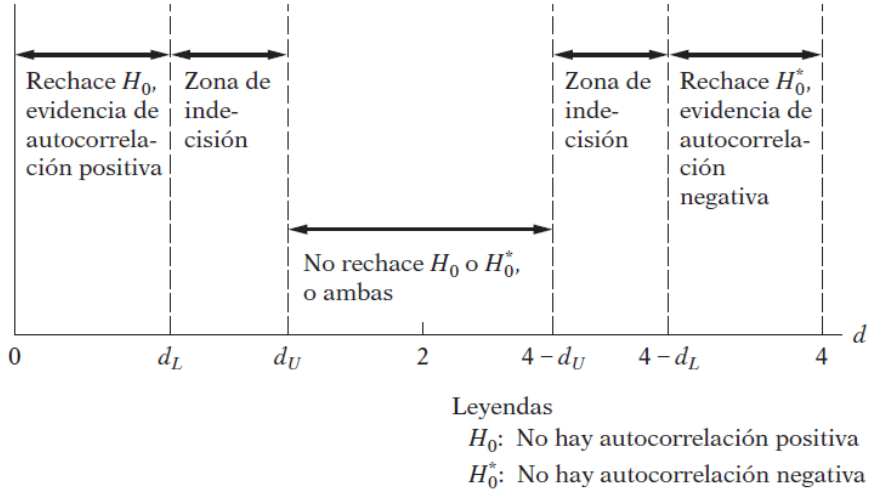
Para la normalidad y homogeneidad de varianza pueden ser utilizadas las pruebas de Shapiro-Wilks modificado y Bartlett respectivamente, las cuales se presentaron previamente en apartado de diagnóstico de modelos ARIMA. Para modelos de regresión es frecuente realizar la prueba de Incorrelación de Durbin Watson como lo expone Aparicio et al. (2004).

El estadístico d de Durbin Watson que es utilizado para probar las hipótesis se define como (Gujarati y Porter, 2010):

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{t} e_i^2} \qquad \text{(Ec. 29)}$$

El estadístico d se prueba comparándolo con los valores críticos d_L y d_U que se determinan en función del número de variables regresoras del modelo y el número de observaciones. La hipótesis nula que se desea probar es H_0: No hay autocorrelación positiva o negativa. El esquema de decisión se muestra en la figura 3.

Figura 3. **Evaluación del estadístico d de Durbin Watson**



Fuente: Gujarati y Porter (2010). *Econometría* (p. 435).

2.2. Hidrología de centrales hidroeléctricas

En el contexto de centrales hidroeléctricas, el caudal entrante es aquel flujo de agua por unidad de tiempo que llega hacia el embalse en el caso de centrales con almacenamiento o el flujo de agua que llega a la toma de agua para conducir a las turbinas en el caso de centrales sin almacenamiento como se lo describe PSR Energy Consulting (2018). Los caudales son variables de entrada de modelos complejos de planificación de la operación de sistemas eléctricos.

2.2.1. Blance hídrico

Aplicar métodos estadísticos a los históricos de caudales de las centrales hidroeléctricas permiten obtener mejores estimaciones y acotar los escenarios extremos posibles y determinar los escenarios con mayor probabilidad, aumentando con ello la probabilidad de incurrir en mejores decisiones del uso de los recursos energéticos del país.

Una variable importante a tener en cuenta es el de caudal total que estará disponible para el llenado del embalse o para el turbinamiento en una central. Esto porque el modelo conceptual de la relación entre caudal de la estación hidrológica y la generación de una central puede diferir cuando se consideren caudales parciales en lugar de caudales totales.

En el modelo de balance hídrico usado por PSR Energy Consulting (2018) el caudal total afluente a una central estará constituido por la suma de los caudales afluentes más los caudales turbinados que aportan las defluencias de otras estaciones aguas arriba.

El balance hidro está asociado espacial y temporalmente para cada central hidroeléctrica de la siguiente manera: el almacenamiento al final de la etapa t (inicio de la etapa t + 1) es equivalente al almacenamiento inicial menos el desfogue total (turbinamiento y vertimiento) más el volumen afluente (caudales adyacentes más el desfogue de las centrales aguas arriba) (PSR Energy Consulting [PSR], 2018).

2.2.2. Límites de almacenamiento y turbinamiento

Cada embalse tiene sus características de almacenamiento y desfogue, niveles máximos y mínimos, que se relacionan en los modelos con los caudales entrantes, para definir la mejor forma de hacer uso del agua disponible en los embalses y el agua entrante. Los caudales tratados en este trabajo básicamente son los caudales Entrantes y Caudales Laterales (PSR Energy Consulting [PSR], 2018).

Los límites de turbinamiento y almacenamiento representan una variable a considerar que limita la producción de energía ante crecidas o disminuciones considerables de caudales afluentes.

2.2.3. Coeficientes de producción

Las centrales hidroeléctricas tienen características de producción expresada normalmente en términos de volumen de agua por unidad de tiempo o caudal entrante turbinado. La potencia de salida de una central hidroeléctrica es proporcional a la magnitud del caudal Q, a la altura del desplazamiento del caudal turbinado h y a la eficiencia de la turbina η, pero suele simplificarse en la solución de problemas de despacho de generación, mediante la obtención de un factor o coeficiente de producción ρ que está en función de h y η, expresando la salida de la central únicamente en términos del caudal turbinado como (PSR Energy Consulting [PSR], 2018):

$$P_g = \rho * Q \qquad \text{(Ec. 30)}$$

Donde:

P_g es la potencia generada por la turbina o central hidroeléctrica.

ρ es el coeficiente de producción de la central.

Q es el caudal entrante turbinado por la central.

Con el uso del factor de producción se pueden realizar estimaciones de forma simplificada sobre el potencial de generación si se tienen pronósticos de caudales afluentes, pero debe tenerse en cuenta la gestión que se hace de los caudales disponibles y diferenciarlo de los caudales turbinados.

3. PRESENTACIÓN DE RESULTADOS

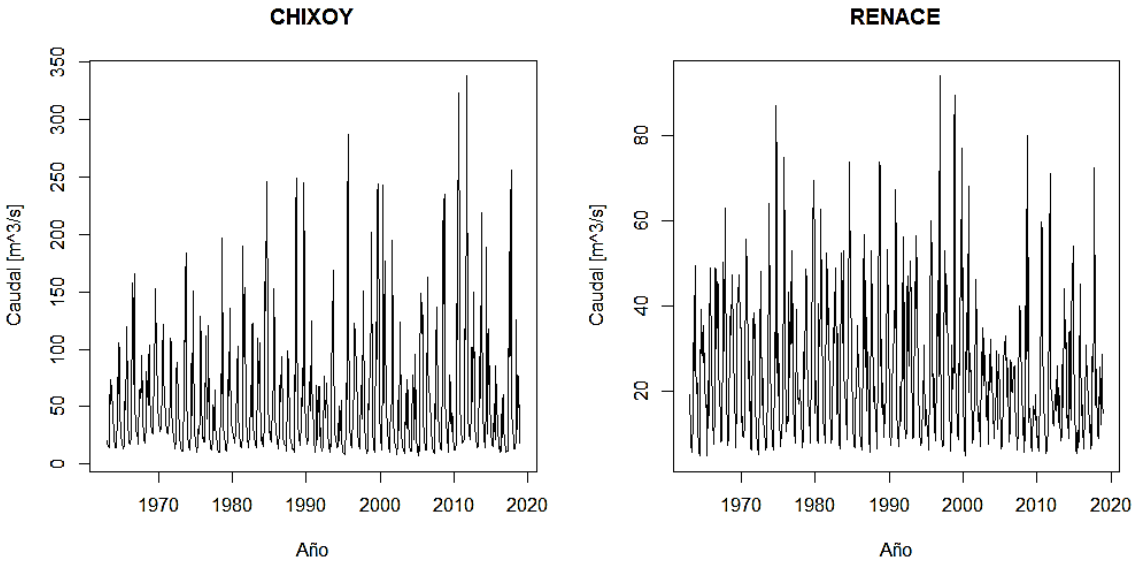
3.1. Metodología de desarrollo y selección de los modelos

Se seleccionaron dos centrales hidroeléctricas de interés para las cuales se desarrolló un modelo estadístico con base en la información disponible de caudales promedio mensuales y energía generada. Por ser dos de las principales hidroeléctricas del país y tener información de caudales para el periodo de 1963 a 2019 e información de energía generada para periodo de 2006 a 2020 para las hidroeléctricas CHIXOY y RENACE, se desarrolló el presente trabajo para estas centrales.

La metodología determinada para el modelo estadístico se compone en una primera etapa de pronóstico de caudales promedios mensuales con modelos SARIMA y una segunda etapa para la estimación de la generación hidroeléctrica por medio de un modelo de regresión lineal por segmentos. La metodología e información adecuada utilizada para desarrollar y seleccionar los modelos de mejor ajuste se describe a continuación.

Para los modelos de pronóstico de caudales se analizaron las series temporales univariadas de los promedios mensuales asociados a cada central hidroeléctrica para el período de 1963 a 2018 de la base de datos de la Programación de Largo Plazo publicada por el Administrador del Mercado Mayorista que se muestran en la figura 4. Estas series no muestran una tendencia de crecimiento o decrecimiento definida, pero si se observa una componente estacional y valores extremos altos que se intensifican en diferentes periodos a lo largo del tiempo.

Figura 4. **Series temporales de caudal promedio mensual 1963-2018**



Fuente: elaboración propia, realizado con RStudio.

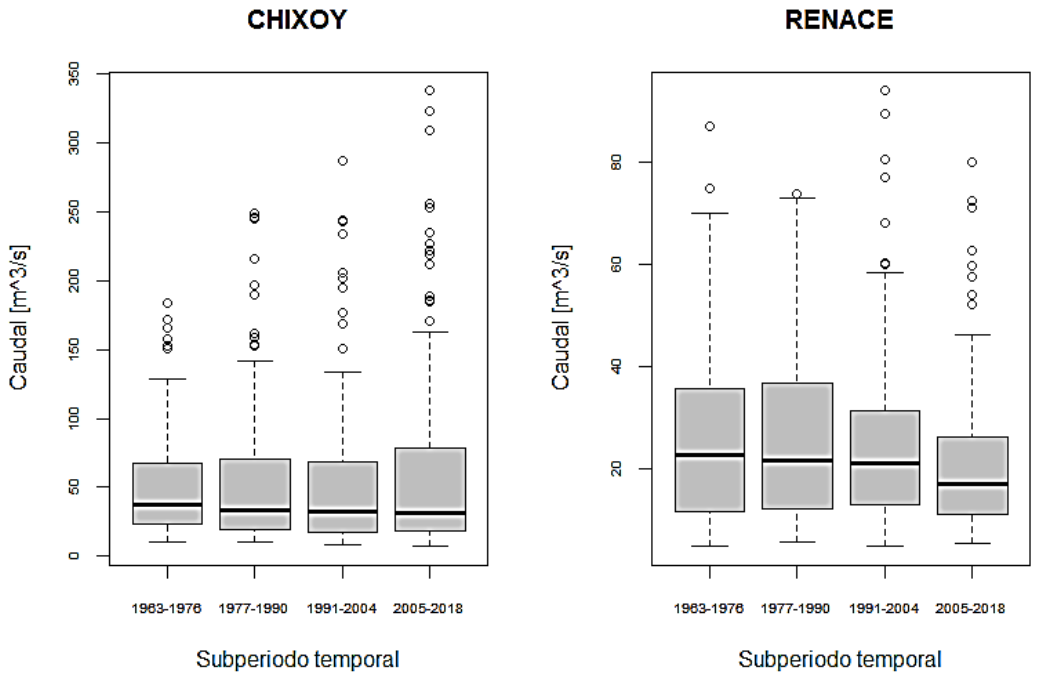
3.1.1. Análisis exploratorio de los caudales promedios mensuales

El análisis exploratorio previo a la metodología Box-Jekins se realizó con base en el trabajo de Meis y Llano (2017) y se complementó con gráficos de la descomposición de la serie con base en el modelo aditivo.

Para visualizar posibles cambios en los valores medios de la serie entre períodos de igual duración se obtuvieron gráficos de caja que se muestran en la figura 5 para cuatro períodos de 14 años 1963-1976, 1977-1990, 1991-2004 y 2005-2018. Para los caudales de CHIXOY se observa un incremento leve entre el primero y segundo período y entre el tercer y cuarto período en el rango intercuartílico y en los valores extremos altos, pero no se tienen cambios notorios para la mediana. Para la serie de RENACE por el contrario se tiene una reducción sostenida en el rango intercuartílico e incremento de valores extremos altos a

partir del tercer periodo, así como una reducción notoria en la mediana para el cuarto período.

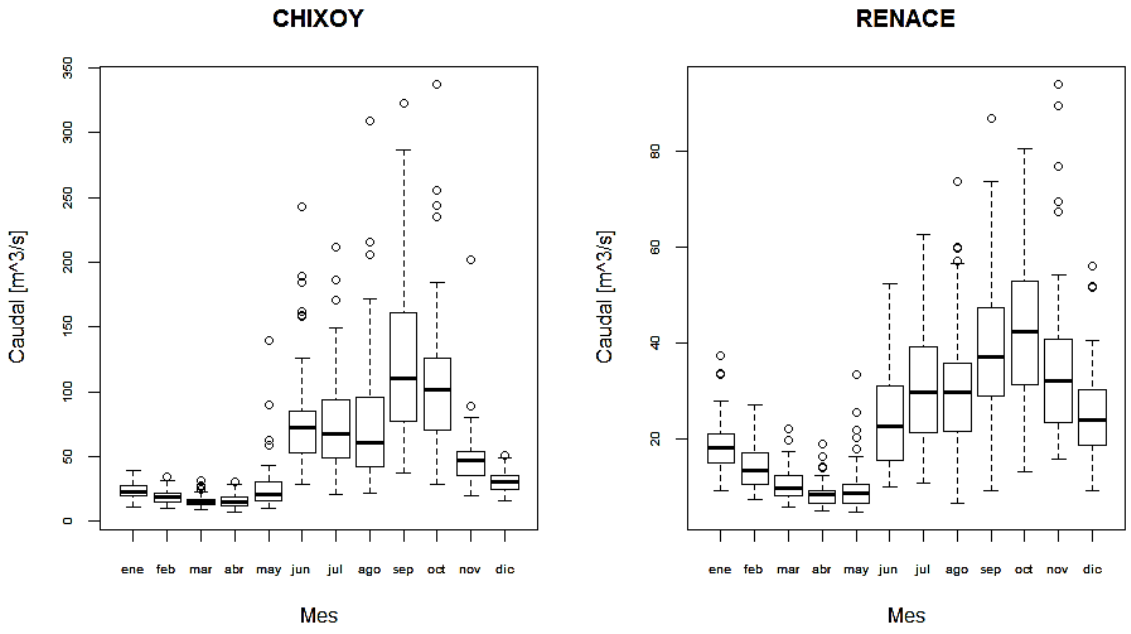
Figura 5. **Boxplot del caudal mensual por subperiodos 1963-2018**



Fuente: elaboración propia, realizado con RStudio.

Para conocer si las series presentaban estacionalidad se realizaron los gráficos de cajas de la figura 6 para datos agrupados por mes. Para la serie de CHIXOY se observa estacionalidad con una onda marcada por valores bajos de caudal promedio mensual de noviembre a mayo y valores altos de junio a octubre, teniendo el máximo en septiembre y presencia de valores extremos altos para el período de valores alto de caudal. La serie de RENACE también muestra estacionalidad con un período de valores bajos de caudal de enero a mayo y el período de valores altos de caudal de junio a diciembre, el caudal máximo se tiene para el período de octubre y se tiene mayor presencia de valores extremos en meses de caudales bajos.

Figura 6. **Boxplot del caudal mensual por mes 1963-2018**

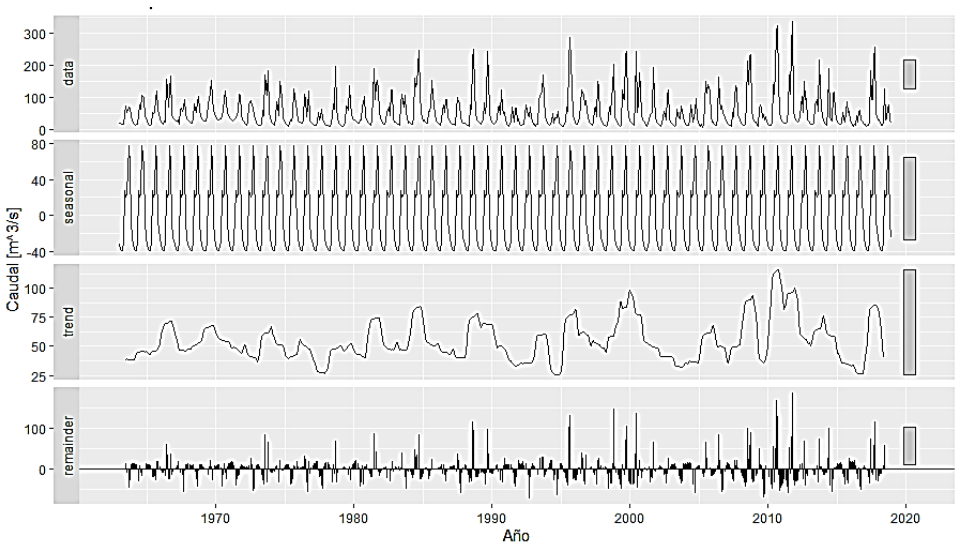


Fuente: elaboración propia, realizado con RStudio.

Finalmente se evaluaron los gráficos de descomposición de la serie de tiempo por medio del modelo aditivo para conocer la tendencia, estacionalidad y variación aleatoria presentes en la serie. La figura 7 muestra la descomposición para la serie de CHIXOY donde puede observarse en la componente de tendencia que se presentan valores superiores a los ocurridos en años anteriores, se muestra evidencia de una componente estacional y variación aleatoria con presencia de valores altos alrededor del año 2010.

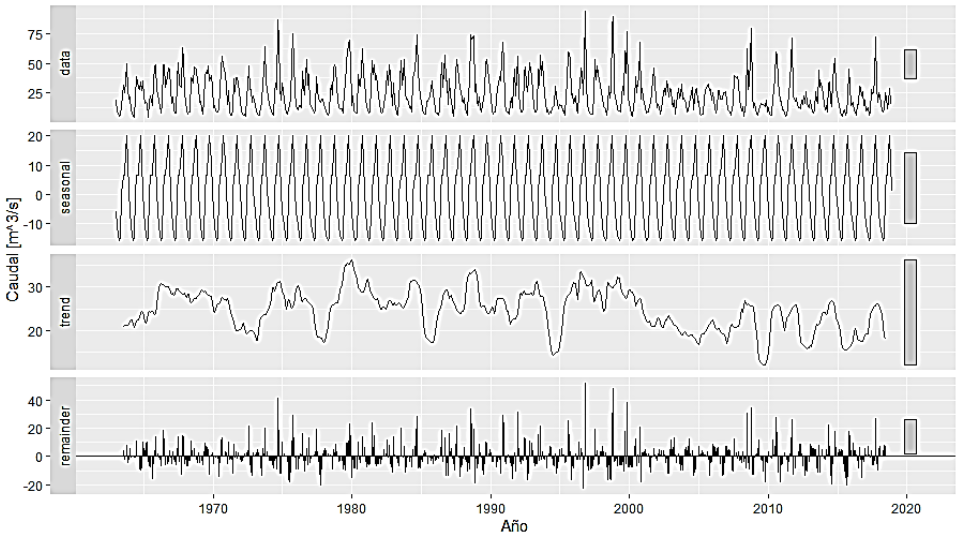
La descomposición de la serie de RENACE de la figura 8 muestra en su componente de tendencia una clara reducción de los valores a partir del año 2004, una observación que confirma lo observado en el gráfico de cajas de la figura 5, también puede observarse la presencia de la componente estacional y una reducción de la componente aleatoria a partir del 2004 a excepción de valores altos alrededor del año 2010.

Figura 7. **Descomposición modelo aditivo CHIXOY**



Fuente: elaboración propia, realizado con realizado con RStudio.

Figura 8. **Descomposición modelo aditivo RENACE**



Fuente: elaboración propia, realizado con RStudio.

3.1.2. Desarrollo y selección de modelos SARIMA

Una vez conocido el comportamiento de la serie de tiempo se aplicó la metodología Box-Jekins para diferentes conjuntos de datos en busca de obtener el modelo de mejor ajuste para los caudales de cada central en el cual su residuo fuera un ruido blanco. Los conjuntos analizados fueron:

- Todos los datos de 1963 a 2018 sin transformaciones.
- Todos los datos de 1963 a 2018 con transformación Box-Cox.
- Todos los datos de 1963 a 2018 con transformación de Johnson.
- Datos para el periodo 2005 a 2018 sin transformaciones.
- Datos para el periodo 2005 a 2018 con transformación Box-Cox.
- Datos para el periodo 2005 a 2018 con transformación de Johnson.

El conjunto de datos que permite que los modelos estimados cumplan con tener residuos de ruido blanco es el periodo 2005 a 2018 con transformación de Johnson. La transformación de Johnson se realizó con base en la ecuación (19) con los siguientes parámetros para la familia SB:

Tabla I. **Parámetros de transformación de Johnson**

	Parámetro			
Serie	**γ**	**λ**	**ε**	**η**
CHIXOY	1.7279	414.9920	6.8188	0.6890
RENACE	2.3983	142.3132	3.0198	1.0980

Fuente: elaboración propia.

El desarrollo de la metodología Box-Jekins para ambas centrales se muestra en los siguientes cuatro pasos:

Paso 1. Identificación del modelo

Inicialmente se evaluó la estacionariedad de las series utilizando la prueba de Dickey-Fuller Aumentada obteniendo los resultados de la tabla II los cuales muestran que para ambas series se tienen valores del estadístico de prueba DF menores a los valores críticos de Dickey-Fuller al 5 %, por lo cual la hipótesis nula de existencia de raíz unitaria se rechaza y por tanto la serie es estacionaria, la misma conclusión se tiene al utilizar el p valor del estadístico de prueba con un nivel de significancia de 0.05.

Tabla II. **Prueba de estacionariedad de las series transformadas**

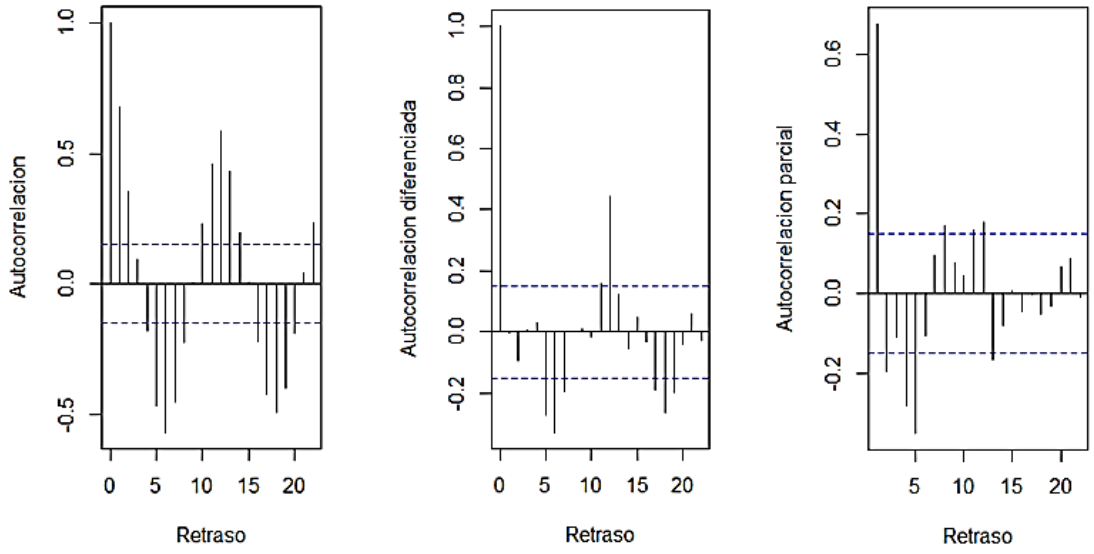
Variable	hipótesis	n	t 5%	DF	p(Unilateral)
Caudal CHIXOY SB	H_0: δ = 0 (no estacionaria) H_1: δ < 0 (estacionaria)	168	-1.95	-8.4782	0.01
Caudal RENACE SB	H_0: δ = 0 (no estacionaria) H_1: δ < 0 (estacionaria)	168	-1.95	-8.6512	0.01

Fuente: elaboración propia.

Para confirmar que se mantiene la estacionalidad de las series originales sin transformaciones e identificación de posibles valores p, d, q, P, D y Q , se evaluaron los gráficos de autocorrelación (ACF) y autocorrelación parcial (PACF). En la figura 9 se tienen los gráficos para CHIXOY donde se observa que 10 de las 12 primeras autocorrelaciones son significativas mostrando algún indicio de no estacionariedad aun cuando esta hipótesis fue rechazada, por lo que se evaluaron las autocorrelaciones en primera diferencia de la serie para evaluar la posibilidad de requerir esta diferencia, pero se observa que las primeras autocorrelaciones no son significativas y se rechaza la posibilidad de evaluar los modelos con $d = 1$. Al observar los gráficos de la figura 6 se observa que la serie transformada presenta estacionalidad por lo que se evaluaron los modelos con

la serie diferencia estacionalmente $D = 1$. El grafico de autocorrelaciones muestra picos significativos en los retardos $q = 1,2,5,6,7$ y el gráfico de autocorrelaciones parciales muestra picos significativos que sugieren retardos significativos $p = 1,2,4,5$.

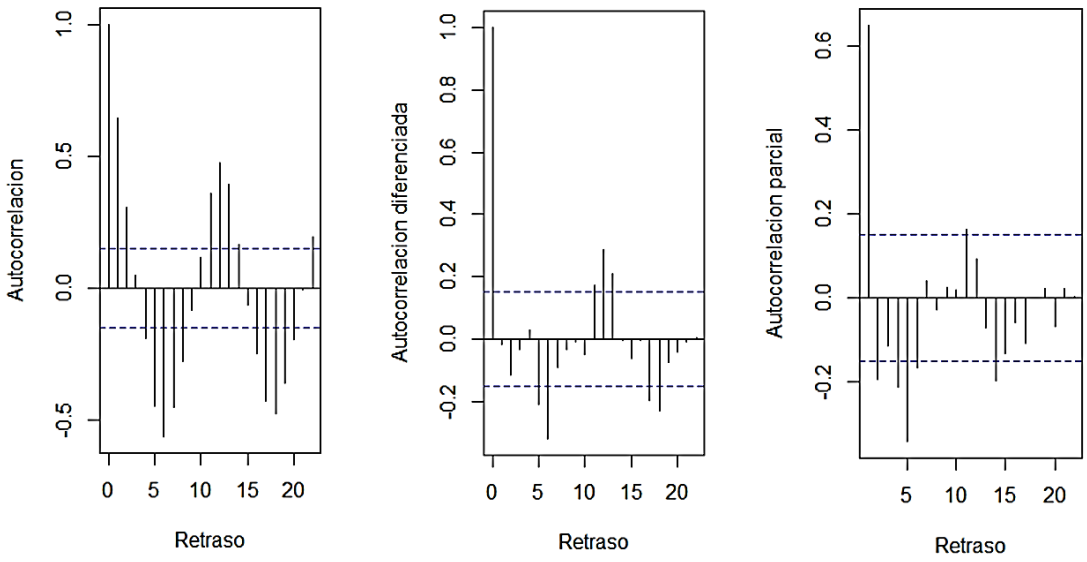
Figura 9. **Autocorrelación y Autocorrelación parcial de CHIXOY**



Fuente: elaboración propia, realizado con RStudio.

Los gráficos correspondientes a RENACE se tienen en la figura 10 donde puede observarse que el gráfico de autocorrelaciones sugiere nuevamente que pueda existir una componente no estacionaria, pero al evaluar el gráfico en primera diferencia se concluye nuevamente que no es necesario analizar los modelos con la serie en primera diferencia de la parte no estacional $d = 1$. La estructura estacional de los gráficos sugiere evaluar los modelos para la serie diferenciada estacionalmente $D = 1$. El gráfico de autocorrelaciones muestra picos significativos en los retardos $q = 1,2,4,5,6,7$ y el grafico de autocorrelaciones parciales muestra picos significativos que sugieren retardos significativos $p = 1,2,4,5$.

Figura 10. **Autocorrelación y Autocorrelación parcial RENACE**



Fuente: elaboración propia, realizado con RStudio.

Paso 2. Estimación de parámetros

Se realizó la estimación de los parámetros utilizando software para las combinaciones p, d, q, P, D, Q , con cantidad de parámetros reducidos para obtener modelos sencillos y evitar el sobreajuste permitiendo evaluar y analizar de esta manera un conjunto reducido de modelos de manera individual. Para CHIXOY se probaron los 81 modelos para $d = 0$, $D = 1$ con p, q , P y Q entre 0 y 2. Para RENACE se probaron los 36 modelos posibles modelos para $d = 0$, $D = 1$ con p, q entre 0 y 2, y P y Q entre 0 y 1. Para todos los modelos se obtuvieron los índices error medio (ME), error cuadrático medio (RMSE), error medio absoluto (MAE), Criterio de Akaike (AIC) y Criterio de información Bayesiano (BIC) para determinar el modelo de mejor ajuste.

Los modelos con mejor ajuste tienen índices similares y en algún caso se tiene un modelo diferente seleccionado para cada índice por lo cual se eligió el modelo que fuera seleccionado por la mayor cantidad de índices. Las tablas III y IV muestran el comparativo de los mejores 5 modelos obtenidos para cada central. Para CHIXOY el mejor modelo se obtiene el modelo SARIMA $(1,0,1)(2,1,1)_{12}$ por ser el mejor para los criterios ME, AIC y BIC puesto que los modelos$(2,0,0)(0,1,1)_{12}$ y $(1,0,1)(0,1,1)_{12}$ no mostraban tener residuos de ruido blanco. Para RENACE el modelo seleccionado fue el SARIMA $(2,0,2)(0,1,1)_{12}$ siendo el mejor para los criterios RMSE, MAE y AIC.

Tabla III. **Modelos de mejor ajuste para CHIXOY**

p	d	q	P	D	Q	ME	RMSE	MAE	AIC	BIC
2	0	0	0	1	1	0.0127	0.5032	0.3602	261.31	273.51
1	0	1	0	1	1	0.0119	0.5028	0.3591	260.93	273.12
1	0	1	2	1	1	0.0050	0.5005	0.3590	262.18	280.48
1	0	1	2	1	2	0.0058	0.5001	0.3584	264.11	285.46
1	0	2	2	1	2	0.0058	0.5001	0.3587	266.11	290.51

Fuente: elaboración propia.

Tabla IV. **Modelos de mejor ajuste para RENACE**

p	d	q	P	D	Q	ME	RMSE	MAE	AIC	BIC
1	0	0	0	1	1	-0.0008	0.5464	0.4233	303.6	312.8
2	0	0	0	1	1	-0.0007	0.5463	0.4232	305.6	317.8
2	0	2	0	1	1	-0.0069	0.5258	0.4074	303.4	321.7
2	0	1	1	1	1	-0.0067	0.5286	0.4091	304.6	322.9
2	0	2	1	1	1	-0.0066	0.5265	0.4074	305.3	326.6

Fuente: elaboración propia.

La tabla V muestra el resumen de los coeficientes con su correspondiente error estándar entre paréntesis para el modelo de mejor ajuste de cada central.

Tabla V. **Parámetros de los modelos de mejor ajuste**

Serie	ar1	ar2	ma1	ma2	sar1	sar2	sma1
CHIXOY SB	0.7661 (0.0917)		-0.1745 (0.1524)		-0.1745 (0.1524)	-0.2126 (0.1231)	-0.6247 (0.1608)
RENACE SB	1.5865 (0.1278)	-0.6420 (0.1238)	-1.2000 (0.1774)	0.2000 (0.1743)			-0.9997 (0.1635)

Fuente: elaboración propia.

Paso 3. Examen diagnóstico de los residuos

Para el modelo de mejor ajuste seleccionado para cada central se realizó el examen diagnóstico utilizando la prueba de Shapiro-Wilks Modificada para la normalidad, la prueba de Bartlett para la igualdad de varianza y la prueba de Box-Pierce para la autocorrelación. El resumen de los resultados se tiene en la tabla VI en la cual se puede observar que los todos los p valor de los estadísticos de pruebas indican que no se pueden rechazar las hipótesis nulas de residuos distribuidos normalmente, igualdad de varianza para dos grupos y significancia en la autocorrelación de los residuos. Debe notarse que la prueba de Shapiro-Wilks es la prueba modificada para muestras con $3 \leq n \leq 5000$. La prueba de Bartlett fue realizada para muestras balanceadas constituidas por la mitad de los residuos separadas con criterio cronológico. Para la prueba de significancia de las autocorrelaciones se muestran los resultados con el estadístico Q evaluado hasta el rezago 10.

Tabla VI. **Diagnóstico de los residuos de los modelos de mejor ajuste**

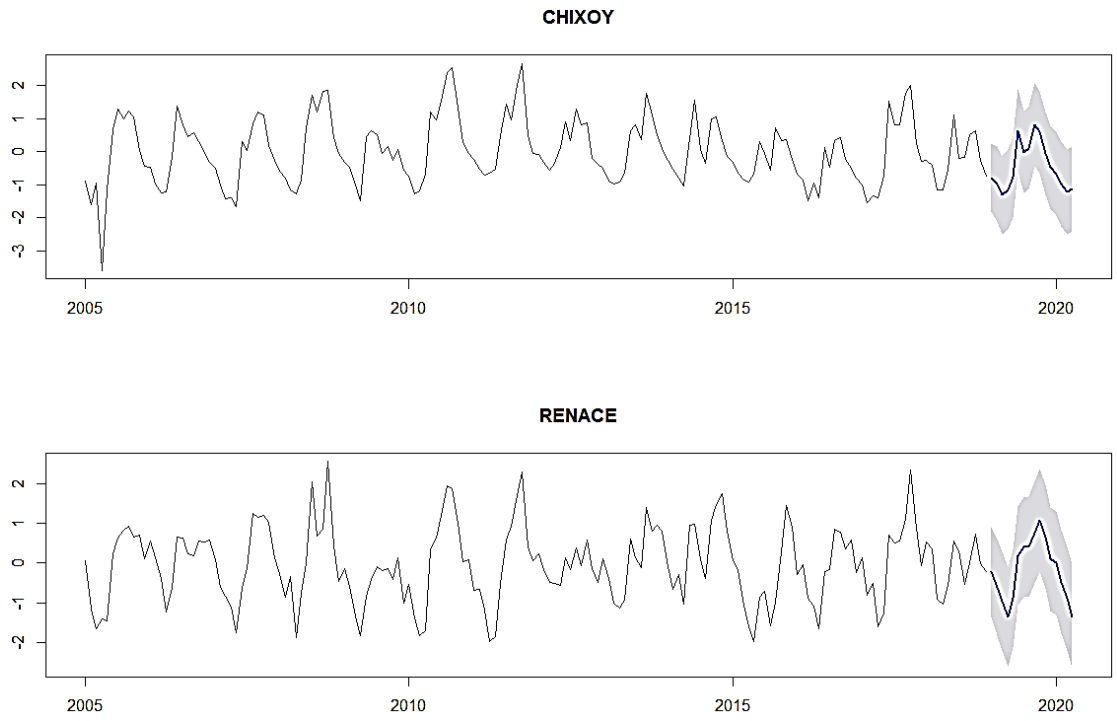
Prueba de Shapiro-Wilks (modificado) para normalidad						
Hipótesis	Serie	n	Media	Desv. Est.	W*	p valor
H0: Distr. normales	RENACE SB	168	-0.0069	0.52736	0.97831	0.1349
H1: Distr. no normales	CHIXOY SB	168	0.0050	0.5020	0.9776	0.1155
Prueba de Bartlett de homogeneidad de varianza						
Hipótesis	Serie	n	Media	Varianza	K	p valor
H0: $\sigma_1^2 = \sigma_2^2 = \sigma^2$	RENACE SB	84	-0.0427	0.2536	0.65969	0.4167
		84	0.0289	0.3033		
H1: $\sigma_1^2 \neq \sigma_2^2 \neq \sigma^2$	CHIXOY SB	84	0.0808	0.2905	2.5092	0.1132
		84	0.0289	0.2048		
Prueba de Box-Pierce de autocorrelación en los residuos						
Hipótesis	Serie	n	Media	gl (k)	Q	p valor
H0: $\rho_1 = \rho_2 = \rho_k = 0$	RENACE SB	156	-0.0069	10	1.8134	0.9752
H1: Algún $\rho_k \neq 0$	CHIXOY SB	156	0.0050	10	1.8134	0.9976

Fuente: elaboración propia.

Paso 4. Pronósticos del modelo

En esta etapa se utilizaron los coeficientes de los modelos para obtener el pronóstico con sus intervalos de confianza obteniendo los resultados que se visualizan en la figura 10. Los pronósticos se realizaron para 16 etapas desde enero 2019 hasta abril 2020. Como se observa en la figura 11 los pronósticos se realizan para la variable transformada donde los coeficientes determinados tienen validez.

Figura 11. **Pronóstico de 16 etapas para variables transformadas**



Fuente: elaboración propia, realizado con RStudio.

En la sección 3.2 se presentan con mayor detalle los resultados de los pronósticos anteriores y se presentan los pronósticos equivalentes transformados en la dimensional requerida por los modelos de regresión de la segunda etapa.

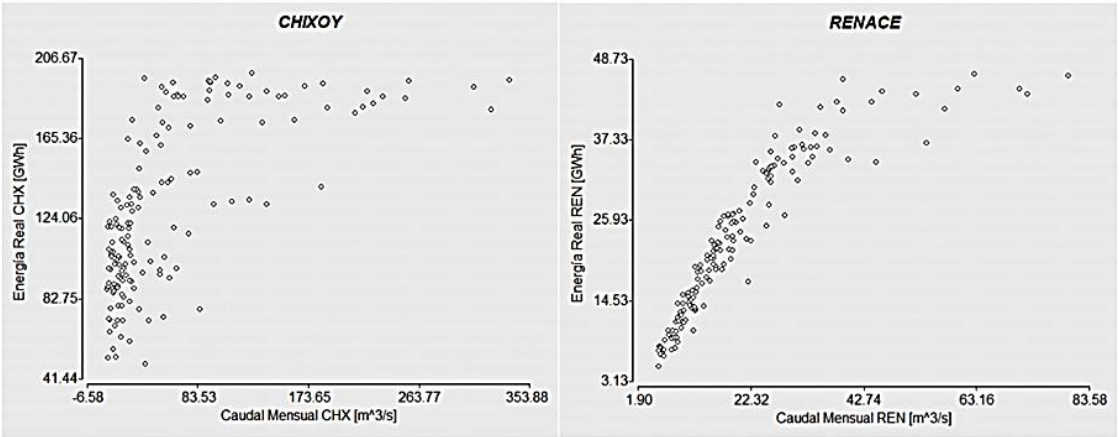
3.1.3. Desarrollo y selección de los modelos de regresión

Para este trabajo se utilizó un modelo de regresión lineal que permite obtener la energía generada con base en el caudal promedio mensual obteniendo con ello un modelo simplificado que explique la variación de los diferentes factores externos al caudal.

Debido a los objetivos del trabajo únicamente se consideró el caudal promedio mensual como variable regresora realizando un procedimiento secuencial para la elección del modelo de mejor ajuste basado en el Coeficiente de Determinación R^2 , la desviación estándar estimada y el cumplimiento de los supuestos de normalidad, homocedasticidad e independencia de los residuos.

Para corroborar la conocida relación lineal entre el caudal y la energía generada se realizó el gráfico de dispersión que muestra la figura 11. Se puede observar una relación lineal con la existencia de valores atípicos entre los valores bajos y un punto a partir del cual la energía generada no parce tener incremento aun con aumentos altos del caudal. Los caudales máximos turbinables de diseño de las centrales son 75 m^3/s para CHIXOY y 34.59 m^3/s para RENACE valores que son coincidentes con los puntos a partir de los cuales se observa un cambio en la relación entre el caudal y la energía generada.

Figura 12. **Dispersión de energía mensual generada 2006-2018**



Fuente: elaboración propia, con datos obtenidos del AMM. *Planificación de la Operación y Resultados de la Operación*. Consultado el 01 de agosto de 2020. Recuperado de: https://www.amm.org.gt.

Para RENACE se probaron inicialmente los modelos de regresión lineal simple y regresión polinómica de orden 2 y orden 3. El modelo de regresión lineal simple brinda un coeficiente R^2 de 0.7880, un ajuste bajo por lo que se probaron otros modelos. Los modelos de orden 2 y orden 3 muestran una mejora notable en el ajuste con R^2 de 0.9372 y 0.9439, respectivamente; sin embargo, el supuesto de normalidad para residuos no se cumple.

En busca de obtener un modelo de mayor ajuste que cumpla en mejor medida los supuestos básicos de los residuos se probó el modelo de regresión lineal por segmentos utilizando como valor umbral el caudal máximo turbinable de diseño de la central. Con el modelo por segmentos se obtuvo un R^2 de 0.9364 que cumple con el examen diagnóstico de sus residuos por lo que fue el modelo seleccionado.

El resumen de los estadísticos de la bondad de ajuste de regresión de los modelos que fueron evaluados para RENACE se muestra en la tabla VII.

Tabla VII. **Bondad de ajuste de regresión para RENACE**

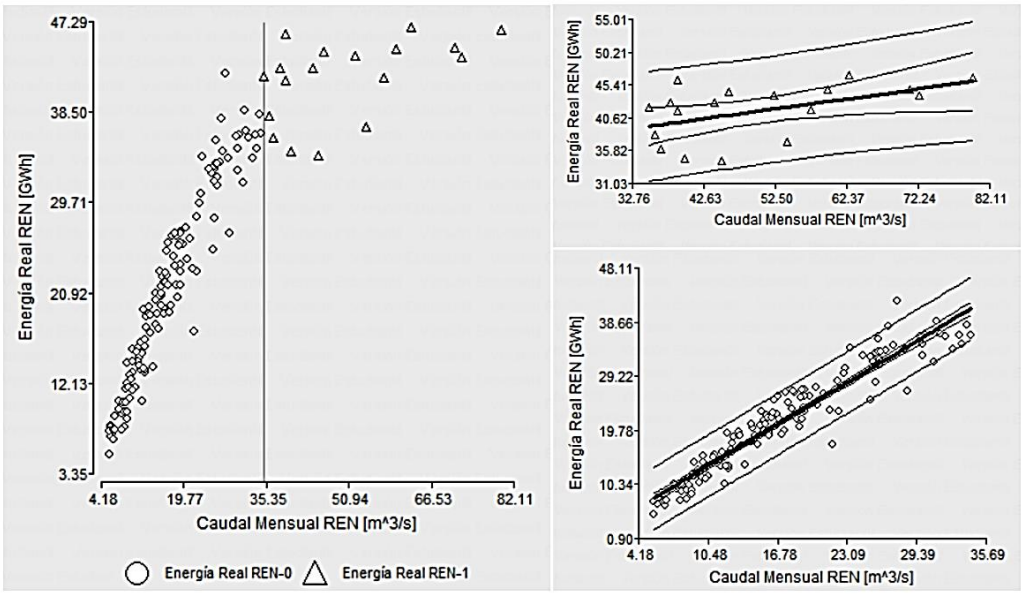
Modelo	R^2	R^2 Aj	Varianza	Desviación Estándar
Orden 1	0.7880	0.7866	26.0928	5.1081
Orden 2	0.9372	0.9364	7.7772	2.7888
Orden 3	0.9439	0.9428	6.9928	2.6444
Segmentos	0.9364	0.9360	7.8312	2.7984

Fuente: elaboración propia.

En la figura 12 se puede observar el ajuste que muestra el modelo seleccionado para cada segmento de la variable dicótoma que toma el valor de

0 para los datos con caudal menor o igual al caudal máximo turbinable de 34.59 m^3/s y valor de 1 para los datos con caudal mayor 34.59 m^3/s.

Figura 13. **Regresión por segmentos RENACE**



Fuente: elaboración propia, con datos obtenidos del AMM. *Planificación de la Operación y Resultados de la Operación*. Consultado el 01 de agosto de 2020. Recuperado de: https://www.amm.org.gt

Para CHIXOY se probaron los mismos modelos de regresión lineal y polinómicas de orden 2 y orden 3 pero para la variable regresora caudal promedio mensual al mes correspondiente de energía generada y conociendo la capacidad de almacenaje de su embalse que permite trasladar energía en períodos mensuales se probaron los modelos para la variable regresora caudal promedio mensual del mes anterior al de la energía generada. El resumen de estadísticos de la bondad de ajuste de los modelos se muestra en la tabla VIII donde puede observarse que para todos los modelos se obtiene un mejor ajuste con la variable del caudal del mes anterior. El mejor ajuste se tiene para el modelo de regresión lineal por segmentos con variable rezagada con un R^2 de 0.6347.

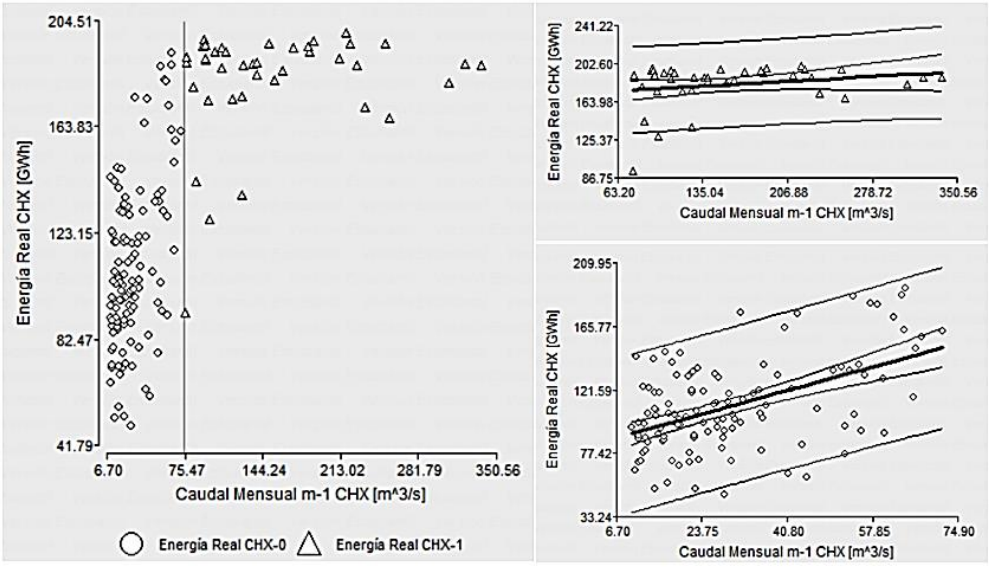
Tabla VIII. **Bondad de ajuste de regresión para CHIXOY**

Modelo	R^2	R^2 Aj	Varianza	Desviación estándar
Orden 1	0.4371	0.4335	1033.4938	32.1480
Orden 2	0.5540	0.5482	824.1871	28.7087
Orden 3	0.5780	0.5697	785.0298	28.0184
Segmentos	0.5783	0.5755	774.3747	27.8276
Orden 1 m-1	0.4800	0.4766	954.7765	30.8995
Orden 2 m-1	0.6206	0.6156	701.2226	26.4806
Orden 3 m-1	0.6282	0.6208	691.7309	26.3008
Segmentos m-1	0.6347	0.6323	670.8158	25.9001

Fuente: elaboración propia.

La figura 14 muestra el ajuste de regresión lineal por segmentos obtenido para CHIXOY donde la variable dicótoma cambia de valor para el umbral 75 m^3/s.

Figura 14. **Regresión por segmentos CHIXOY**



Fuente: elaboración propia, con datos obtenidos del AMM. *Planificación de la Operación y Resultados de la Operación*. Consultado el 01 de agosto de 2020. Recuperado de: https://www.amm.org.gt.

En la etapa de diagnóstico de los residuos de los modelos se verificó la normalidad de los residuos por medio de la prueba de Shapiro-Wilks modificado, para la igualdad de varianzas con la prueba de Bartlett y para la independencia se utilizó la prueba de Durbin Watson

El resumen de los estadísticos de prueba e hipótesis evaluadas se tiene en la tabla 15 donde puede determinarse con base en el p valor de los estadísticos de prueba W* y K que se no se pueden rechazar las hipótesis de normalidad e igualdad de varianza entre los grupos 0 y 1 evaluados. La hipótesis de independencia de los residuos si se rechaza por tener un estadístico d menor a $dL\ y\ dU$ ubicándolo en la zona de rechazo que se mostro en la figura 3.

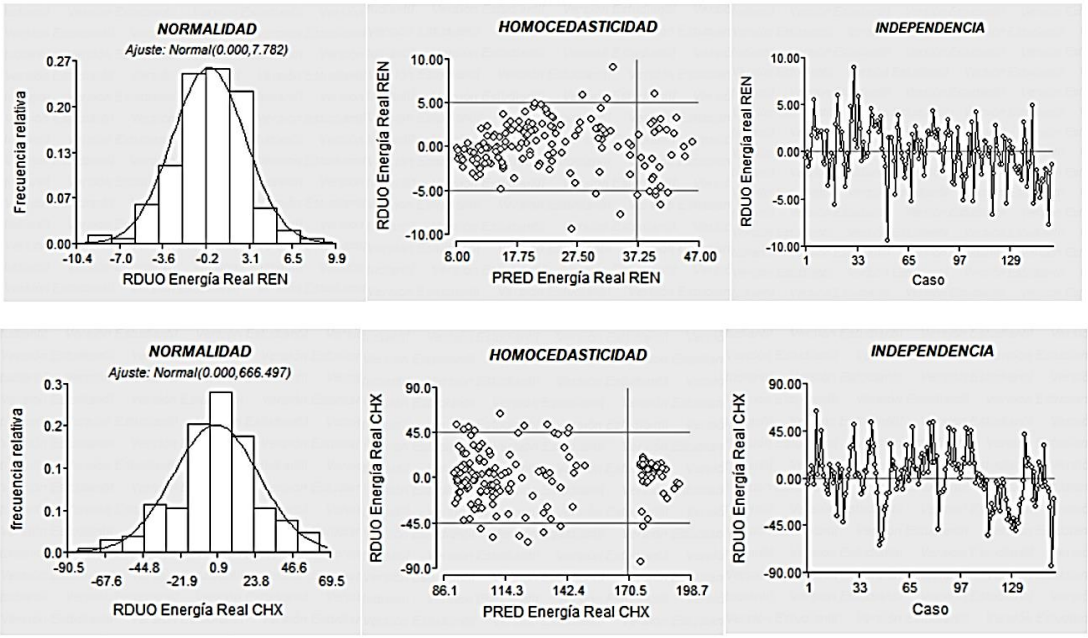
Tabla IX. **Evaluación de supuestos básicos del modelo de regresión**

Prueba de Shapiro-Wilks (modificado) para normalidad						
Hipótesis	Central	n	Media	Desv. Est.	W*	p valor
H0: Distribución normal H1: Distribución no normal	RENACE	156	0.0000	2.7896	0.9899	0.8940
	CHIXOY	156	0.0000	25.8166	0.9773	0.1466
Prueba de Bartlett de homogeneidad de varianza						
Hipótesis	Central	n	Media	Varianza	K	p valor
H0: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ H1: $\sigma_1^2 \neq \sigma_2^2 \neq \sigma^2$	RENACE	138	0.0000	7.3092	2.1108	0.1463
		18	0.0000	12.0499		
	CHIXOY	116	0.0000	751.1599	3.9449	0.04701
		40	0.0000	433.9401		
Prueba de incorrelación de Durbin Watson						
Hipótesis	Central	n	Media	d	dL	dU
H0: No hay autocorrelación positiva H1: Hay autocorrelación positiva	RENACE	156	0.0000	1.395	1.72	1.746
	CHIXOY	156	0.0000	0.8272	1.72	1.746

Fuente: elaboración propia.

La evaluación gráfica de los residuos de los modelos se puede observar en la figura 15 que contiene el histograma para la normalidad, la dispersión de predichos respecto de los residuos para a homocedasticidad y la serie de residuo por caso parala independencia.

Figura 15. **Supuestos de los residuos de regresión**



Fuente: elaboración propia, realizado con InfoStat.

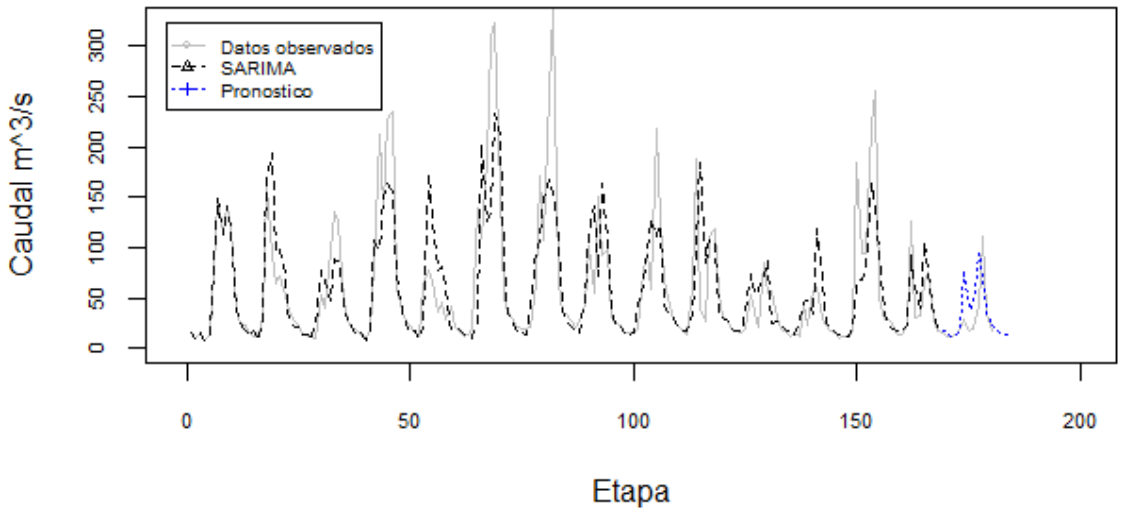
3.2. Confiabilidad de los modelos para escenarios hidrológicos

Los pronósticos generados para las variables transformadas en el paso 4 de la metodología Box-Jenkins fueron transformados a sus unidades originales para ser utilizados como variable de entrada de los modelos de regresión que permiten cuantificar la producción hidroeléctrica asociada. Para obtener los valores originales se despejó de la ecuación (19) de la familia SB de

transformación de Johnson y utilizando los mismos parámetros mostrados en la tabla I.

En la figura 16 se muestra la simulación del modelo en sus variables originales con ajuste y pronóstico de caudales para el escenario medio para CHIXOY. Se observa que el ajuste da un seguimiento adecuado a la estacionalidad y presenta poca diferencia en los meses de bajos caudales de noviembre a mayo y para los meses de caudales altos en los cuales no se sobrepasan los 125 m^3/s, para los años en los que se presentan valores de caudales arriba de los 150 m^3/s la diferencia es notoria entre ajuste y los valores reales.

Figura 16. **Ajuste y pronóstico del caudal de CHIXOY**

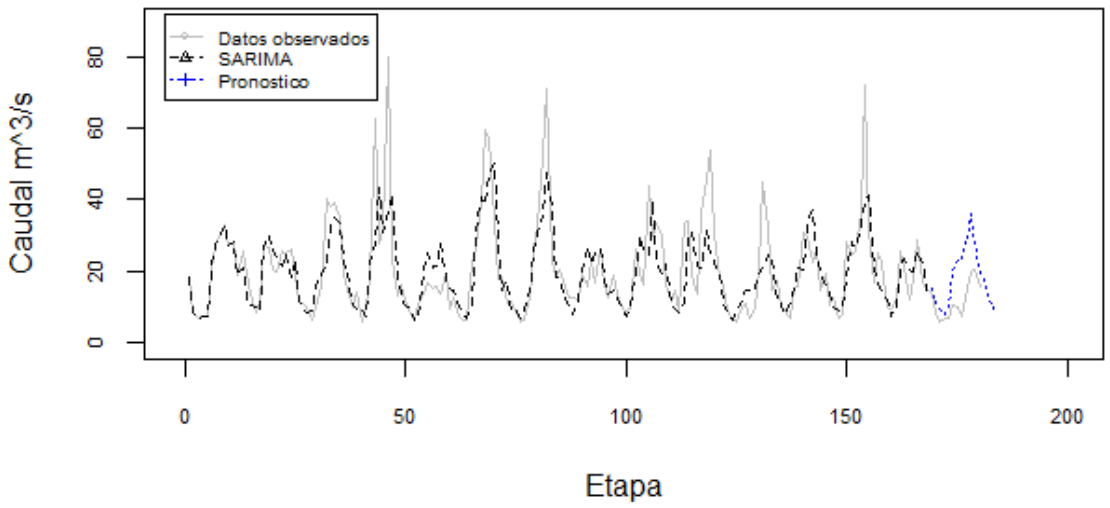


Fuente: elaboración propia, realizado con RStudio.

El caudal simulado para el modelo SARIMA de RENACE en variables originales se muestra en la figura 17 donde puede observarse un comportamiento similar al presentad por el modelo CHIXOY. Para valores reales arriba de los 35

m^3/s se observan diferencias importantes en el ajuste del modelo dado que se tienen subestimaciones del caudal.

Figura 17. **Ajuste y pronóstico del caudal de RENACE**



Fuente: elaboración propia, realizado con RStudio.

Para los resultados mostrados en las dos gráficas anteriores se obtuvieron las métricas Error Medio Absoluto (MAE) y Error Absoluto Porcentual Medio (MAPE) para el conjunto de datos de ajuste y para el conjunto de datos pronosticados para evaluar el nivel de confiabilidad del modelo. Para los datos de ajuste se obtuvieron valores del MAE de 23.81 m^3/s y 5.64 m^3/s para CHIXOY y RENACE respectivamente y valores del MAPE de 36.71 % para CHOCOY y 26.16 % para RENECE. Para los datos pronosticados se obtuvieron valores del MAE de 16.52 m^3/s y 7.69 m^3/s así como valores de MAPE de 55.93 % y 72.00% para CHIXOY y RENACE respectivamente. Estos índices se resumen en la tabla X.

Tabla X. **Índices de los modelos en variables originales**

Indice	CHIXOY	RENACE
MAE Ajuste	23.81	5.64
MAE Pronóstico	16.52	7.69
MAPE Ajuste	36.71%	26.16%
MAPE Pronóstico	55.93%	72.0%

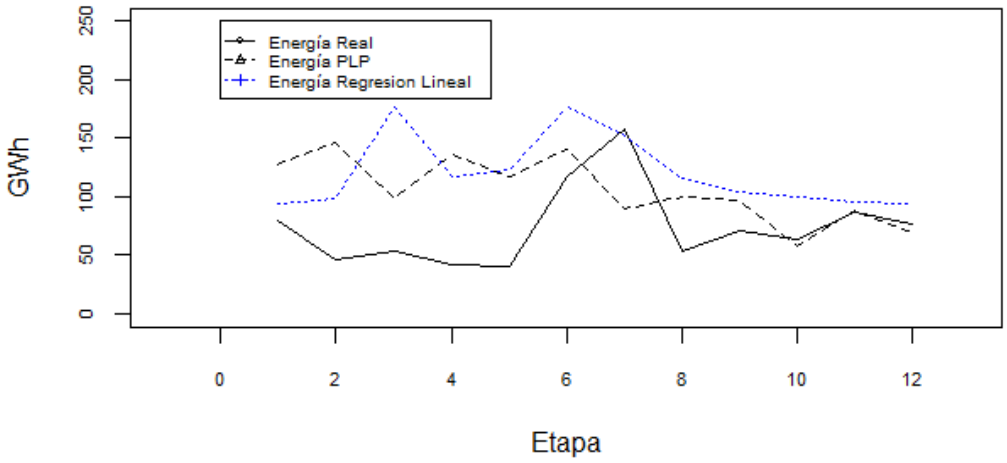
Fuente: elaboración propia.

3.3. Comparación de escenarios pronosticados producción hidroeléctrica

Para cuantificar la mejora en la producción hidroeléctrica se utilizó la segunda etapa de los modelos de mejor ajuste constituido por los modelos de regresión lineal por segmentos para generar los pronósticos de generación entre mayo 2019 y abril 2020 y se compararon los resultados con el escenario de generación hidroeléctrica que presenta AMM (2019) en la Programación de Largo Plazo para las centrales analizadas.

La comparación de escenarios pronosticados para CHIXOY puede observarse gráficamente en la figura 18 que muestra los pronósticos de generación y la energía real generada. El pronóstico generado con el modelo de regresión lineal por segmentos muestra un ajuste apropiado a las variaciones que se presentan en el período evaluado, pero las magnitudes presentadas evidencian diferencias considerables en la estimación de la energía. En cuatro de los 12 períodos los valores pronosticados por el modelo de regresión presentan mayor certeza que la estimación realizada en la PLP.

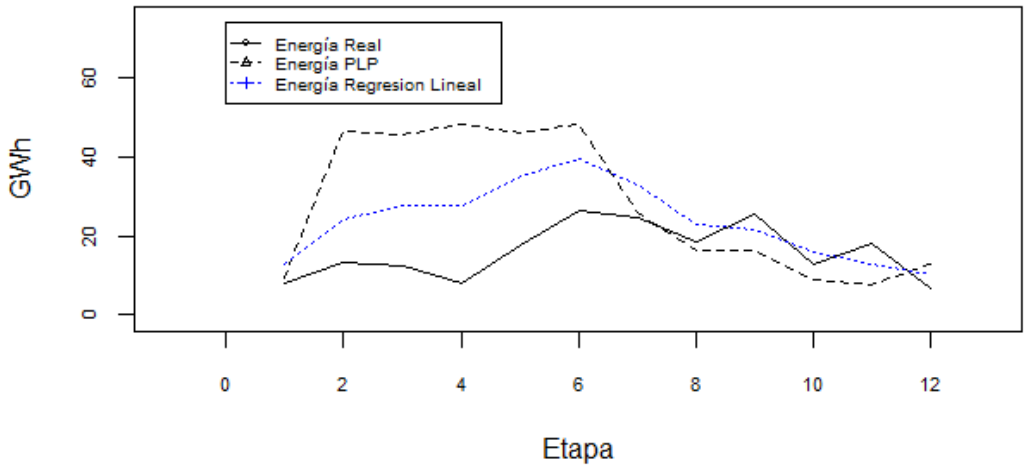
Figura 18. **Comparación de escenarios pronosticados para CHIXOY**



Fuente: elaboración propia, con datos obtenidos del AMM. *Planificación de la Operación y Resultados de la Operación*. Consultado el 01 de agosto de 2020. Recuperado de: https://www.amm.org.gt.

La comparación de escenarios pronosticados para RENCE se tienen en la figura 19. La mejora en el pronóstico con el modelo de regresión lineal por segmentos es notable respecto a al pronóstico PLP para 8 de los 12 periodos analizados y en los restantes 4 períodos se observa un ajuste adecuado a la energía real generada.

Figura 19. **Comparación de escenarios pronosticados para RENACE**



Fuente: elaboración propia, con datos obtenidos del AMM. *Planificación de la Operación y Resultados de la Operación*. Consultado el 01 de agosto de 2020. Recuperado de: https://www.amm.org.gt.

Para cuantificar la mejora en la producción hidroeléctrica se utilizó el índice del Valor Medio Absoluto (MAE) porque brinda un comparativo en las unidades de interés. Para CHIXOY los valores del MAE son 45.15 GWh y 47.67 GWh para el pronóstico PLP y el modelo de regresión por segmentos respectivamente, lo que representa un incremento del 6% en la incerteza del pronóstico de la producción hidroeléctrica. Para RENACE se obtuvieron valores del MAE de 15.92 GWh para el pronóstico PLP y 9.18 GWh para el pronóstico de regresión lineal por segmentos teniendo con ello una reducción de la incerteza del 42 %.

4. DISCUSIÓN DE RESULTADOS

4.1. Análisis del desarrollo y selección de los modelos

En la práctica de la planificación de los sistemas eléctricos se utilizan pronósticos de caudales promedios mensuales como variable de entrada a modelos de planificación como lo describen Morales (2016) y Zuñiga y Jordán (2005), por lo cual un modelo de pronóstico de caudales representa únicamente una parte de un modelo que permita determinar escenarios de generación hidroeléctrica.

En este trabajo no se emplea un modelo de planificación para determinar la generación hidroeléctrica asociada a los caudales y se propone el uso de un modelo estadístico de regresión para obtener los valores de energía de las centrales bajo análisis generando de esta manera las dos etapas del modelo estadístico que se presentaron en los resultados.

Morales (2016) menciona la importancia de tomar un período de 30 años en el estudio de caudales para representar fenómenos climáticos y de comportamiento de largo plazo por lo cual se consideró para el análisis exploratorio de las series de caudales el período completo de información disponible desde 1963 hasta 2018, teniendo un total de 56 años.

En el análisis exploratorio se evaluaron aspectos como la tendencia, estacionalidad, variaciones en la mediana y rangos intercuartílicos de subperíodos de análisis y presencia de valores atípicos como los sugiere Meis y Llano (2017) comprobando que el análisis realizado permite conocer los aspectos

relevantes del comportamiento de las series de caudales identificando la presencia de componentes de la serie que permiten conocer preliminarmente la estructura que tendrán los modelos SARIMA y obtener subperiodos de análisis que permiten modelar la serie para períodos determinados en casos de cambios significativos en los estimadores como la media y varianza de las series.

Del análisis exploratorio se tuvieron indicios de posibles problemas en el diagnóstico de los modelos SARIMA generados para ambas series de tiempo, debido a la presencia de valores atípicos que se incrementaban en el subperiodo 2005-2018 y en los meses de altos caudales a lo largo del año como se presentó en las figuras 5 y 6. También se observó que existían indicios de cambios de cambios en la media y varianza de la serie a partir del año 2005 debido a los cambios en el rango intercuartílico y la mediana del gráfico de cajas de la figura 5 así como en la componente de tendencia del modelo aditivo de las figuras 7 y 8.

En el apartado de resultados de desarrollo y selección de modelos SARIMA se listó los diferentes grupos de datos para los cuales se aplicó la metodología Box-Jekins. Para los 6 conjuntos de datos se tuvieron inconvenientes en la normalidad de los residuos de todos los modelos probados para los tres conjuntos que utilizaban la serie completa de 1963-2018 y para dos conjuntos de datos con la serie para el periodo 2005-2018. El único conjunto de datos que brinda varios modelos con normalidad en los residuos es para la serie del periodo 2005-2018 con transformación de Johnson. La posibilidad de usar únicamente una parte de la serie para obtener un modelo adecuado se basa en los resultados mostrados por Meis y Llano (2016) que sugieren la posibilidad de usar los subperíodos de la serie para modelarla correctamente puesto que en ocasiones no se tienen buenos resultados con un solo modelo SARIMA.

La cantidad de datos del período sobre el cual se desarrollaron los modelos finales se reduce a un cuarto de los datos de la serie completa por lo que se debe esperar que los residuos de los modelos presenten normalidad.

En la tabla I se mostraron los resultados de la prueba de Dickey-Fuller donde se comprobó que ambas series son estacionaras por lo cual no es necesario que la series sean diferenciadas y por tanto no es necesario incluir en el análisis de mejores modelos los que contienen $d = 1$.

Para verificar este hecho se realizó el análisis gráfico empleado por Meis y Llano (2017) al evaluar los gráficos de autocorrelaciones de las series diferenciadas cuando el grafico de autocorrelaciones muestra varios rezagos significativos, pero con los resultados mostrados en las figuras 9 y 10 se puede observar que lo primeros rezagos no son significativos en la serie diferenciada.

En la etapa de identificación de posibles valores p, d, q, P, D y Q con base en los gráficos de autocorrelación y autocorrelación parcial se mostró que existían rezagos significativos adicionales a los empleados en los modelos evaluados para seleccionar el de mejor ajuste.

El uso de modelos con un número reducido de coeficientes empleado en el trabajo responde a la necesidad de obtener modelos simples capaces de modelar adecuadamente la serie de tiempo y evitar sobre ajustes innecesarios del modelo.

Los modelos finales fueron seleccionados con base en los criterios ME, RMSE, MAE, AIC y BIC debido a que en los estudios previos revisados no se tiene un criterio único o preferido para la selección de modelos. Como se presentó en los resultados, el modelo de mejor ajuste de cada serie se seleccionó con

base en varios criterios, pero ambos modelos seleccionados comparten la característica de poseer el mejor MAE y AIC que son los criterios utilizados por Dmitrieva (2015), Zuñiga y Jordán (2005) y Meis y Llano (2016).

La tabla V de los modelos de mejor ajuste muestra que para el caso de CHIXOY la serie se modela adecuadamente con tres coeficientes que corresponden a la parte estacional y dos coeficientes que corresponden a la parte no estacional mientras que para RENACE se tienen cuatro coeficientes para la parte no estacional y un solo coeficiente para la parte estacional.

También se observa que CHIXOY posee tres coeficientes para el proceso autorregresivo y dos coeficientes para el proceso de medias móviles mientras que RENACE tiene dos coeficientes para el proceso autorregresivo y tres coeficientes para el proceso de medias móviles.

La conformación de los modelos de mejor ajuste basada en sus coeficientes indica que tanto para CHIXOY como para RENACE se tienen modelos explicados con la misma cantidad de coeficientes para la componente de valores pasados y componente del ruido para la parte no estacional.

Mientras que para la parte estacional CHIXOY es explicado por una cantidad mayor de retardos de valores pasados que por la componente de ruido y para la componente estacional de RENACE únicamente se tiene una aportación de la componente de ruido.

Las pruebas utilizadas para el diagnóstico de los modelos difieren respecto de algunas pruebas empleadas en trabajos de pronóstico de caudales. Para el caso de la normalidad, Vega (2016) sugiere el uso de técnicas como Anderson Darling o Kolmogorv-Smirnov pero en este trabajo se utilizó la prueba de Shapiro-

Wilks modificado considerando el trabajo específico de Razali y Wah (2011) por tener un enfoque estadístico. Para la prueba de normalidad Vega (2016) y Meis y Llano (2017) sugiere el uso de la prueba e Ljun-Box pero en este trabajo se usó la prueba de Box-Pierce considerando que Gujarati y Porter (2010) indican que son pruebas similares pero la eficiencia de la prueba de Box-Pierce es mayor para muestras grandes.

Respecto a los pronósticos generados con el modelo debe recordarse que la aplicación correcta es realizarlos para la variable transformada SB puesto que los coeficientes fueron determinados para esta esta variable. Debe evitarse obtener un modelo equivalente en las variables originales para realizar pronósticos.

Para el desarrollo del modelo de regresión se asumió la relación lineal entre el caudal y la energía generada con base en la relación que muestra la ecuación (30) y únicamente se verificó gráficamente. No se determinó coeficientes de correlación puesto que para su correcta aplicación las variables deben tener una distribución por lo que hubiese sido necesario realizar transformación de la variable y desarrollar todo el análisis de regresión en variables trasformadas obteniendo un modelo y resultados menos intuitivos. Considerando, que para el análisis de regresión solo debe cumplirse el supuesto de normalidad para los residuales, la aplicación del modelo de regresión en las variables originales es adecuada.

Los modelos de regresión de mejor ajuste fueron los modelos de regresión lineal por segmentos considerando el valor del coeficiente de determinación R^2 y el cumplimiento de los supuestos de los residuales. Para ambos modelos es necesario conocer el valor umbral de la variable regresora que provoca el cambio en la estructura de la regresión como lo indican Gujarati y Porter (2010).

Para este trabajo se utilizó el valor del caudal turbinable máximo de diseño de las centrales, el cual incluye algún error debido a que este caudal está definido para un momento determinado de operación de la central por lo cual no debe ser el mismo valor para la variable de caudal promedio mensual, sin embargo, como se mostró en las figuras 13 y 14 los valores elegidos permiten segmentar adecuadamente los valores observados de energía generada respecto de los caudales.

La aplicación del modelo de regresión lineal por segmentos se justifica en este trabajo por el enfoque de modelar todos los datos observados con un modelo estadístico confiable, pero en la práctica puede evaluarse la posibilidad de limitar un modelo para valores limitados por el caudal turbinable máximo.

Esto sería la parte del modelo con la variable dicótoma $D_i = 0$ equivalente a la regresión que se obtendría omitiendo los valores de caudales que arriba del caudal máximo mensual.

La bondad de ajuste del modelo de CHIXOY con R^2 de 0.6347 y RENACE con R^2 de 0.9364 se encuentran entre el rango de 0.6 y 1 que implican que el modelo explica adecuadamente la varianza. Sin embargo, el valor de CHIXOY se encuentra cerca del límite inferior y aunque estadísticamente pueda ser un modelo válido puede presentar errores importantes al realizar la estimación de la energía generada con base en los caudales.

El inconveniente en el supuesto de independencia de los residuos que se mostró en la tabla IX y la figura 15 puede resolverse al aleatorizar los datos sobre los cuales se aplica la regresión con lo cual se obtienen los estadísticos $d = 2.0790$ y $d = 1.8076$ para RENACE y CHIXOY, respectivamente, con lo cual no se pude rechazar la hipótesis de independencia de los residuos.

Con la verificación anterior se puede observar que la no independencia de los residuos del modelo proviene de la naturaleza de los datos por ser una serie de tiempo y no por especificación del modelo por tal motivo se toma como válido y de mejor ajuste el modelo de regresión lineal por segmentos. El hallazgo de no independencia de los residuos sugiere evaluar otras técnicas para el análisis de series de tiempo que incorporen la dependencia de variables rezagadas.

4.2. Análisis de la confiabilidad de los escenarios hidrológicos

Los pronósticos de escenarios hidrológicos presentados en este trabajo corresponden a la serie de datos esperados del caudal promedio mensual obtenido a través de transformar los datos pronosticados con los modelos SARIMA para las variables de caudales con la transformación de Johnson para la familia SB.

Los resultados muestran que los modelos tienen un seguimiento adecuado de la estacionalidad de los datos y alto ajuste para los niveles de caudales bajos. Para ambos modelos se observa que no son capaces de dar seguimiento a los valores máximos en época de caudales altos, pero esto ocurre para caudales arriba de 100 m^3/s para CHIXOY y 35 m^3/s para RENACE para los años que presentan caudales bajos para todos los meses.

Considerando, los caudales turbinables máximos, así como los modelos de regresión que fueron utilizados esto no representa inconvenientes porque a partir de valores de 75 m^3/s y 34.59 m^3/s par CHIXOY y RENACE respectivamente, las variaciones de la energía generada son mínimas ante altos niveles de caudales por lo que el modelo resulta ser confiable para su aplicación en pronósticos de generación hidroeléctrica.

Los modelos fueron desarrollados para series estacionarias por lo que se esperaría que las etapas pronosticadas de las series formaran parte de esta misma serie estacionaria, sin embargo, para ambas series de caudales se observa la presencia de valores bajos totalmente atípicos para los meses que normalmente presentan caudales altos.

Con estos hallazgos podría pensarse que los modelos SARIMA son menos efectivos para estimar valores extremos bajos en comparación con otros modelos que incorporan información climática, pero la comparación de resultados de la energía eléctrica que se discuten más adelante muestra que la previsión de caudales basados en modelos climáticos era más alta que los obtenidos por los modelos SARIMA.

Los índices MAE y MAPE para los valores de ajuste y valores pronosticados de la variable original fueron determinados para obtener un nivel de confiabilidad para los modelos. Los valores del MAE sugieren que los pronósticos generados tienen un buen ajuste porque se reduce 7.29 m^3/s para el caso de CHIXOY y aumenta solo 2.25 m^3/s para RENACE.

4.3. Análisis de la comparación de escenarios pronosticados

La segunda etapa de pronóstico fue desarrollada como una herramienta estadística que reemplaza los modelos de planificación de un sistema eléctrico los cuales obtienen estimaciones de energía eléctrica asociada los caudales proyectados con base en ecuaciones o funciones basadas en la ecuación (30).

Una primera opción para obtener la energía eléctrica asociada a los caudales estimados por los modelos SARIMA podría pensarse con el uso directo

de la ecuación (30), pero para un uso adecuado es necesario que el factor de producción sea constante para todo el rango de caudales y que la energía eléctrica sea determinada en un instante o periodo para el cual se conoce el valor del coeficiente de producción ρ.

Usar un ρ promedio constante implica un error importante debido a que el caudal a evaluar es promedio mensual por lo cual podría obtenerse únicamente una potencia promedio mensual y no la energía asociada. Para obtener la energía asociada al caudal promedio mensual se tendría que estimar un número de horas de operación con la condición de caudal promedio mensual, obteniendo con ello un modelo complejo que puede generar distorsión en la energía generada.

Por lo expuesto anteriormente se considera apropiado utilizar un modelo de regresión lineal por segmentos con el mejor ajuste para obtener la energía asociada, porque permite estimar de una manera sencilla valores de energía eléctrica mensual incorporando además la variabilidad que han aportado diferentes factores operativos en la energía eléctrica para diferentes niveles de caudales que se presenten.

Debe considerarse que el modelo de regresión lineal tuvo una especificación adecuada para todos los supuestos de los residuos al considerar una aleatorización de los datos por lo que éste no explica la varianza que puedan presentar factores operativos estaciones en la relación del caudal promedio mensual y la energía eléctrica asociada.

El bajo ajuste con un con un coeficiente de determinación R^2 de 0.6347 para el modelo de regresión de CHIXOY podría considerarse como la principal fuente del error presentado en el pronóstico de la energía hidroeléctrica, pero al compararlo con el error para el modelo de RENACE se observa que la sobreestimación está presente para ambos modelos en las mismas etapas por lo

que se puede concluir que gran parte del error está dado por el pronóstico de caudales de los modelos SARIMA.

Los resultados del pronóstico de generación hidroeléctrica para CHIXOY muestran que el modelo de regresión para la variable del caudal rezagada en una etapa presenta un mejor seguimiento de las variaciones intermensuales de la generación hidroeléctrica respecto de los valores pronosticados PLP.

Este hallazgo resulta importante porque sugiere que la variación en la generación hidroeléctrica real de CHIXOY no está explicada adecuadamente por los criterios que incorpora el modelo de planificación.

Los resultados de CHIXOY muestran un seguimiento adecuado de la variación intermensual de la generación hidroeléctrica de los dos pronósticos generados, pero con una mejora notable en las 6 primeras etapas de pronósticos a pesar de tener diferencias significativas en los pronósticos de caudales. A partir de los resultados es posible concluir que los pronósticos de caudales con los que se estimó la energía PLP fueron superiores a obtenidos con los modelos SARIMA.

Al evaluar la reducción de la incerteza de los pronósticos de RENACE se observa que la metodología empleada en este trabajo es adecuada para realizar pronósticos de centrales que no tienen la capacidad de almacenar los aportes de los caudales afluentes entre periodos mensuales.

La aplicación de la metodología empleada para centrales como CHIXOY no muestran evidencia de mejora en el pronóstico de la energía hidroeléctrica generada basada en los caudales promedios mensuales.

Debido a que los pronósticos de energía son desarrollados por una combinación de dos modelos diferentes, el error presente en el pronóstico estará compuesto por una parte de cada modelo, por lo cual puede considerarse en el uso de series multivariadas como una alternativa que pueden brindar mejores pronósticos si los modelos son desarrollados con el ajuste y especificación adecuada.

CONCLUSIONES

1. Los pronósticos de caudales promedios mensuales con los modelos SARIMA $(1,0,1)(2,1,1)_{12}$ y $(2,0,2)(0,1,1)_{12}$ utilizados como variable de entrada a un modelo de regresión lineal, permiten seleccionar escenarios de producción hidroeléctrica teniendo un incremento del 6% en la incerteza para la central CHIXOY y una reducción del 42 % en la incerteza para la central RENACE.

2. El modelo estadístico de mejor ajuste para una serie de tiempo de caudales promedios mensuales de las centrales CHIXOY y RENACE se obtiene con un período de 14 años con la aplicación de la metodología Box-Jekins para la serie con transformación de Johnson utilizando como criterio de selección los índices MAE, RMSE, AIC y BIC.

3. El modelo de mejor ajuste para selección de escenarios hidrológicos tiene valores del MAE de 23.81 m^3/s y 5.64 m^3/s para CHIXOY y RENACE. Para los datos pronosticados se obtuvieron valores del MAE de 16.52 m^3/s para CHIXOY y 7.69 m^3/s para RENACE.

4. El pronóstico de energía para CHIXOY con el modelo de mejor ajuste presenta un MAE en 47.67 GWh un valor superior en 6% al MAE del pronóstico PLP. El modelo de mejor ajuste para RENACE muestra una alta reducción respecto de la incerteza del pronóstico PLP con MAE de 9.18 GWh que es 42% menor al MAE de 15.92 GWh del pronóstico PLP.

RECOMENDACIONES

1. Para el pronóstico de energía generada por centrales hidroeléctricas con capacidad de embalse a nivel mensual deben evaluarse otro tipo de modelos estadísticos que permitan incorporar variables explicativas adicionales a los caudales promedios mensuales.

2. Considerar el uso de modelos para series de tiempo multivariadas o modelos de regresión multivariados como alternativas para pronósticos de generación hidroeléctrica.

3. Para el desarrollo de modelos SARIMA de series temporales de caudales promedios mensuales de gran cantidad de datos obtener modelos para subperíodos definidos para obtener modelos con mejores índices.

4. Evaluar diferentes conjuntos de datos para obtener modelos para series de tiempo que cumplan con tener un ruido blanco lo cual permite obtener pronósticos confiables.

REFERENCIAS

1. Administrador del Mercado Mayorista. (2019). *Programación de Largo Plazo Versión Definitiva Mayo 2019-Abril 2020*. Recuperado de: https://www.amm.org.gt/pdfs2/programas_despacho/03_PROGRAMAS_DE_LARGO_PLAZO/2019-2020/02_PLP20190501_VD.pdf

2. Aparicio, J., Martínez, A., y Morales, J. (2004). *Modelos Lineales Aplicados en R.* Universidad Miguel Hernández.

3. Ballou, R. (2004). *Logística, Administración de la cadena de suministro*. Quinta edición. México: Thomson.

4. Cárdenas, M., Agudelo, S., Tabares, J. y Velásquez, C. (2014). *Métodos de pronósticos clásicos y bayesianos con aplicaciones.* Universidad de Valencia.

5. Clement, A. (2015). *Regional Forecasting of Inflow and Generation for Small Hydropower Plant (tesis de maestría).* Norwegian University of Science and Technology. Noruega. Recuperado de: https://www.sintef.no/contentassets/5dc6c523a0b647ea9d9a1db819f8c8cf/3b_4_killingtveit.pdf

6. Díaz, A. y Guevara, E. (2016). *Modelación estocástica de los caudales medios anuales en la cuenca del rio Santa, Perú.* Universidad Nacional Agraria La molina.

7. Dmitrieva, K. (2015). *Forecasting of a hidropower plant energy production (tesis de maestría).* Odtfold University College. Noruega. Recuperado de: https://pdfs.semanticscholar.org/6922/d6b5cb0b390e58fa82e80c725cbb197a7c25.pdf

8. Gujarati, D. y Porter, D. (2010). *Econometría*. Quinta edición. México: McGraw Hill, Inc.

9. Hillier, F. y Lieberman, G. (2010). *Introducción a la Investigación de Operaciones*. Novena edición. México: McGraw Hill, Inc.

10. Lagos, I. y Varga, J. (2003). *Sistema de familias de distribuciones de Johnson, una alternativa para el manejo de datos no normales en cartas de control.* Revista Colombiana de Estadística.

11. Mahibbur, R. y Govindarajulu, Z. (1997). *A modification of the test of Shapiro and Wilks for normality.* Journal of Aplied Statistsics.

12. Meis, M. y Llano, M. (2017). *Modelado estadístico del caudal mensual en la baja Cuenta del Plata.* Universidad de Buenos Aires. Argentina. Recuperado de: http://www.meteorologica.org.ar/wp-content/uploads/2018/12/Meis.pdf

13. Morales Y. (2016). *Evaluación y modelación de información hidrológica para propuesta de mejoras en la programación de largo plazo de centrales hidroeléctricas en Chile*. Universidad Chile.

14. Moreno J. y Salazar J. (2008). *Generación de series sintéticas de caudales usando un Modelo Matalas con medias condicionadas*. Universidad Nacional de Colombia.

15. PSR Energy Consulting. (2018). *SDDP, Manual de metodología*. Versión 15.0. Brasil: PSR, 2014. 43 p.

16. Razali, N. y Wah, Y. (2011). *Power comparisons of Shapiro-Wilk, Kilmogorov-Smirnov, Lilliefors and Anderson-Darling Tests.* Universiti Teknlogi MARA.

17. Smith, R., Vélez, J., Velásquez, J., Ceballos, A., Correa L., Góez, C.,…Zapata E. (2004). *Modelos de predicción de caudales mensuales para el sector eléctrico colombiano.* Universidad Nacional de Colombia. Colombia. Recuperado de: http://bdigital.unal.edu.co/6107/1/No._11-2004-3.pdf

18. Vega, J. (2016). *Generación de series sintéticas de recursos renovables variables para estudios de operación y planificación de sistemas eléctricos.* Universidad de Chile. Chile. Recuperado de: http://repositorio.uchile.cl/bitstream/handle/2250/140413/Generacion-de-series-sinteticas-de-recursos-renovables-variables-para-estudios-de-operacion-y-planificacion-de-sistemas-electricos.pdf?sequence=1&isAllowed=y

19. Walpole, R., Myers, R., Myers, S. y Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias.* México: Pearson Educación.

20. Webster, L. (2001). *Estadística aplicada a los negocios y la economía*. Tercera edición. México: McGraw Hill, Inc.

21. Zuñiga, A. y Jordán, C. (2005). *Pronóstico de caudales medios mensuales empleando Sistemas Neurofuzzy.* Revista Tecnológica de la Escuela Superior Politécnica del Litoral.

ANEXOS

Anexo 1. **Tabla de valores críticos de la distribución Durbin Watson**

	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$		$k' = 6$		$k' = 7$		$k' = 8$		$k' = 9$		$k' = 10$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0.610	1.400	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
7	0.700	1.356	0.467	1.896	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
8	0.763	1.332	0.559	1.777	0.368	2.287	—	—	—	—	—	—	—	—	—	—	—	—	—	—
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	—	—	—	—	—	—	—	—	—	—	—	—
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	—	—	—	—	—	—	—	—	—	—
11	0.927	1.324	0.658	1.604	0.595	1.928	0.444	2.283	0.316	2.645	0.203	3.005	—	—	—	—	—	—	—	—
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.379	2.506	0.268	2.832	0.171	3.149	—	—	—	—	—	—
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.445	2.390	0.328	2.692	0.230	2.985	0.147	3.266	—	—	—	—
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	—	—
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.472	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.257	0.502	2.461	0.407	2.667	0.321	2.873	0.244	3.073
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.692	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.732	2.124	0.637	2.290	0.547	2.460	0.461	2.633	0.380	2.806
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.734
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.751	2.174	0.666	2.318	0.584	2.464	0.506	2.613
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.012	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.958	0.874	2.071	0.798	2.188	0.723	2.309	0.650	2.431
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278	0.682	2.396
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251	0.712	2.363
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920	0.950	2.018	0.879	2.120	0.810	2.226	0.741	2.333
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909	0.972	2.004	0.904	2.102	0.836	2.203	0.769	2.306
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900	0.994	1.991	0.927	2.085	0.861	2.181	0.795	2.281
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.080	1.891	1.015	1.979	0.950	2.069	0.885	2.162	0.821	2.257
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884	1.034	1.967	0.971	2.054	0.908	2.144	0.845	2.236
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.877	1.053	1.957	0.991	2.041	0.930	2.127	0.868	2.216
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870	1.071	1.948	1.011	2.029	0.951	2.112	0.891	2.198
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864	1.088	1.939	1.029	2.017	0.970	2.098	0.912	2.180
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859	1.104	1.932	1.047	2.007	0.990	2.085	0.932	2.164
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854	1.120	1.924	1.064	1.997	1.008	2.072	0.952	2.149
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835	1.189	1.895	1.139	1.958	1.089	2.022	1.038	2.088
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1.201	1.930	1.156	1.986	1.110	2.044
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1.212	1.959	1.170	2.010
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939	1.222	1.984
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923	1.266	1.964
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1.837	1.369	1.873	1.337	1.910	1.305	1.948
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901	1.339	1.935
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893	1.369	1.925
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886	1.396	1.916
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881	1.420	1.909
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877	1.442	1.903
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874	1.462	1.898
150	1.720	1.746	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.847	1.608	1.862	1.594	1.877
200	1.758	1.778	1.748	1.789	1.738	1.799	1.728	1.810	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863	1.665	1.874

Fuente: Gujarati y Porter (2010). *Econometría* (p. 888).