



Modelamiento y análisis del comportamiento de la variable: Generación de Energía Eléctrica, en el sector eléctrico peruano utilizando la metodología Box and Jenkins, para la predicción de este recurso.



Modeling and analysis of the behavior of the variable: Electric Power Generation, in the Peruvian electric sector using the Box and Jenkins methodology, for the prediction of this resource.

Américo A. Rubín de Celis Vidal
Escuela Profesional de Ingeniería de Sistemas e Informática
Universidad Nacional de Moquegua

Abstract — In this work, different ARIMA structures are analyzed through the Box and Jenkins methodology, which allow to establish an adequate model and with an optimal performance that allows to make the conservative prediction of the Electric Power Generation resource for the Peruvian electric sector, with the purpose to quantify and establish the trend and future requirement of electrical infrastructure to serve the growing Peruvian electricity market.

Keywords — *Electric Energy, Prediction, ARIMA, Auto Correlation.*

Resumen — En este trabajo son analizadas diferentes estructuras ARIMA, a través de la metodología Box and Jenkins, que permitan establecer un modelo adecuado y con un óptimo desempeño que permitan realizar la predicción conservadora del recurso Generación de Energía Eléctrica para el sector eléctrico peruano, con la finalidad de cuantificar y establecer la tendencia y futuro requerimiento de infraestructura eléctrica para atender el creciente mercado eléctrico peruano.

Palabras clave — *Energía Eléctrica, Predicción, ARIMA, Auto Correlación.*

I. INTRODUCCIÓN Y DESCRIPCIÓN DEL PROBLEMA

De acuerdo a informaciones y datos estadísticos suministrados por el Ministerio de Energía y Minas (MEM), y el Instituto Nacional de Estadística e Informática (INEI) del Perú, es posible establecer la existencia de una preocupación latente respecto al equilibrio entre oferta y demanda de energía eléctrica en el sector eléctrico peruano, la misma que es clave en la implementación de futuros proyectos relacionados con el uso de este recurso, y con él la inversión y desarrollo de políticas estratégicas para el crecimiento económico del país. Del otro lado se encuentran los inversionistas en el rubro energético y dueños de diversas empresas que dependen de

este recurso, entre las cuales se encuentran las empresas mineras que se constituyen como los principales consumidores de este recurso y generadores de divisas para el estado peruano. Es importante mencionar que, analistas en este rubro energético, citados por la Agencia de noticias Reuters (Vélez (2012). <https://lta.reuters.com/article/topNews/idLTASIE81K03M20120221?sp=true>), consideran que el Perú podría sufrir en los años venideros una crisis energética, como lo ocurrido en Sudáfrica el 2008 y Chile el año 2011. Considerando la actual coyuntura política por lo que viene atravesando el país el actual gobierno busca impulsar el crecimiento económico a través de inversiones en el ámbito privado y considera el Sector Eléctrico Peruano como pieza clave para lograr tal propósito, ya que cuenta con una cartera de proyectos en el ámbito minero por 58,500 millones de dólares, para los siguientes años, de acuerdo a información difundida por Europapress (Mazo (2018). <http://www.europapress.es/internacional/noticia-vizcarra-asegura-no-impondra-proyectos-mineros-comunidades-no-deseen-20180419015100.html>). Al respecto, es importante mencionar a Carlos Herrera Descalzi, Ex-ministro de Energía y Minas del Perú y Ex-Decano del Colegio de Ingenieros del Perú, señaló el año 2013 de forma puntual que el Perú se encuentra en peligro por la falta de un programa estratégico principal en el ámbito energético.

En tal sentido, considerando los antecedentes citados anteriormente, estos nos permiten prestar especial atención a la evolución y comportamiento de las variables relacionadas con la Generación y Demanda de este recurso energético. En este trabajo aborda y enfatiza un análisis univariado (variable única) relacionado al comportamiento de la Variable Generación de Energía Eléctrica (VGEE), perteneciente al sector eléctrico peruano. Datos estadísticos discretos recopilados por el MEM a lo largo de 20 años (1995 – 2016), fueron compilados y publicados en un documento denominado Evolución de Indicadores del Sector Eléctrico 1995-2016 (*Ministe-*

rio de Energía y Minas- Perú, *Evolución de Indicadores del Sector Eléctrico 1995-2016*. Dirección General de Electricidad, Dirección de Estudios y Promoción Eléctrica, (2017)), el cual servirá de base para realizar el análisis en este trabajo. Para ello se busca establecer el modelamiento de la variable VGEE utilizando la metodología Box and Jenkins (ampliamente utilizada en series temporales), utilizando las siguientes métricas: Función Auto correlación (FAC), Función Auto correlación Parcial (FACP), Criterio de Información Akaike (AIC), Criterio de Información Bayesiano (BIC) y diferentes tipos de Errores Medios y porcentuales en predicción. Los datos analizados correspondientes a la variable VGEE, poseen un número finito de parámetros, que permite ser representados por los Modelos ARIMA (Auto Regresivos Integrados de Medias Móviles), frecuentemente empleados en la teoría estadística.

De esta manera, el presente trabajo tiene por objetivo establecer un modelo ARIMA adecuado que permita realizar la predicción y comportamiento de la variable Generación de Energía Eléctrica (VGEE), en términos cuantitativos con la finalidad de establecer el impacto futuro en la demanda de este recurso y el incremento en infraestructura e inversión para los siguientes cinco años en el ámbito energético en

nuestro país.

El presente trabajo fue estructurado de la siguiente manera: I) Introducción y descripción del problema, II) El Sector eléctrico peruano, III) Modelamiento: Metodología Box and Jenkins, IV) Resultados y Análisis, y V) Conclusiones.

II. EL SECTOR ELÉCTRICO PERUANO.

Considerando el ámbito nacional, es posible establecer que la mayor parte de actividades económicas dependen directamente del consumo eléctrico, desde empresas mineras, de manufactura, construcción, hasta llegar a los usuarios residenciales que representan el mayor volumen de consumidores. En la Figura 1, se observa el comportamiento histórico correspondiente al número final de consumidores por Sector Económico, en el que se puede apreciar claramente un incremento sostenido de los usuarios Residenciales, proporcional al crecimiento demográfico del país, como se puede ver en la Figura 2. En tanto que los usuarios Industriales y Comerciales tienen crecimientos sostenidos, pero no con la misma velocidad y tendencia correspondientes al de los clientes Residenciales.

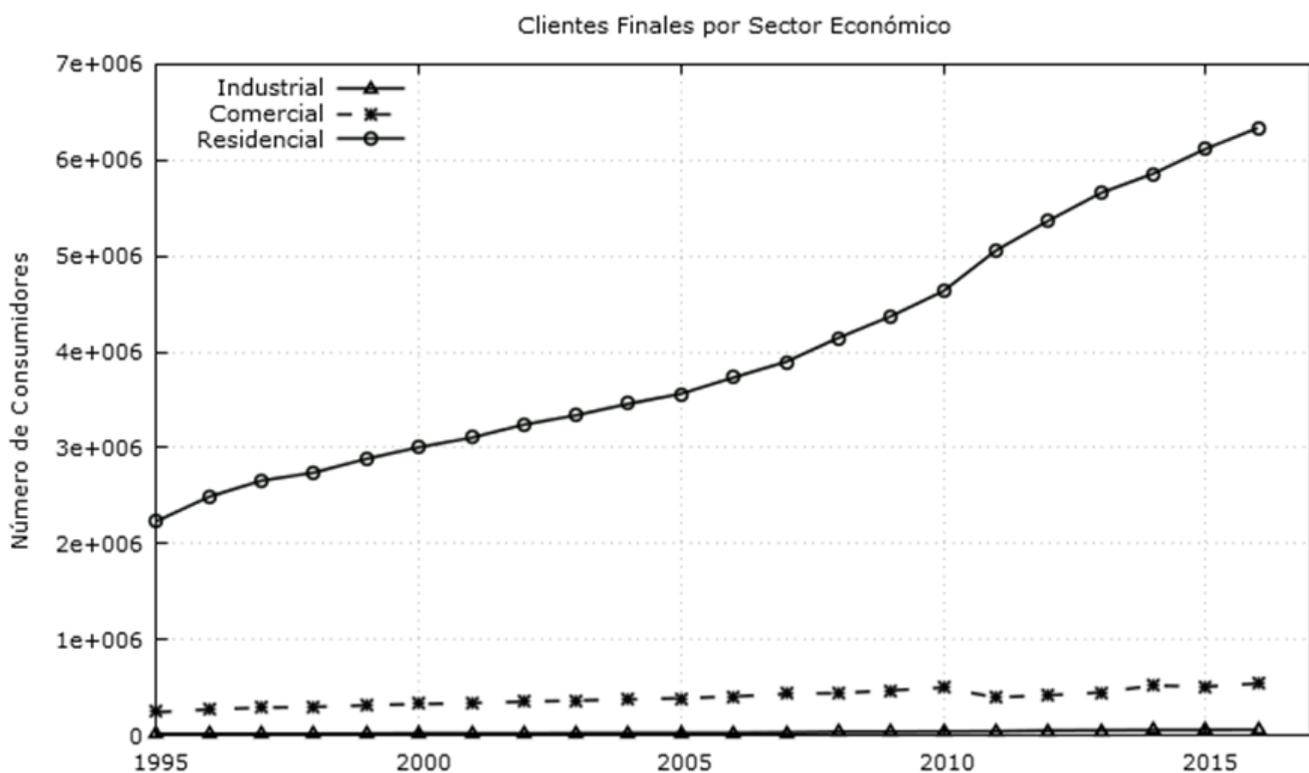


Fig. 1. Número de clientes finales por sector económico, correspondiente al sector eléctrico peruano. Fuente: *Evolución de indicadores del Sector Eléctrico 1995-2016*, MEM.

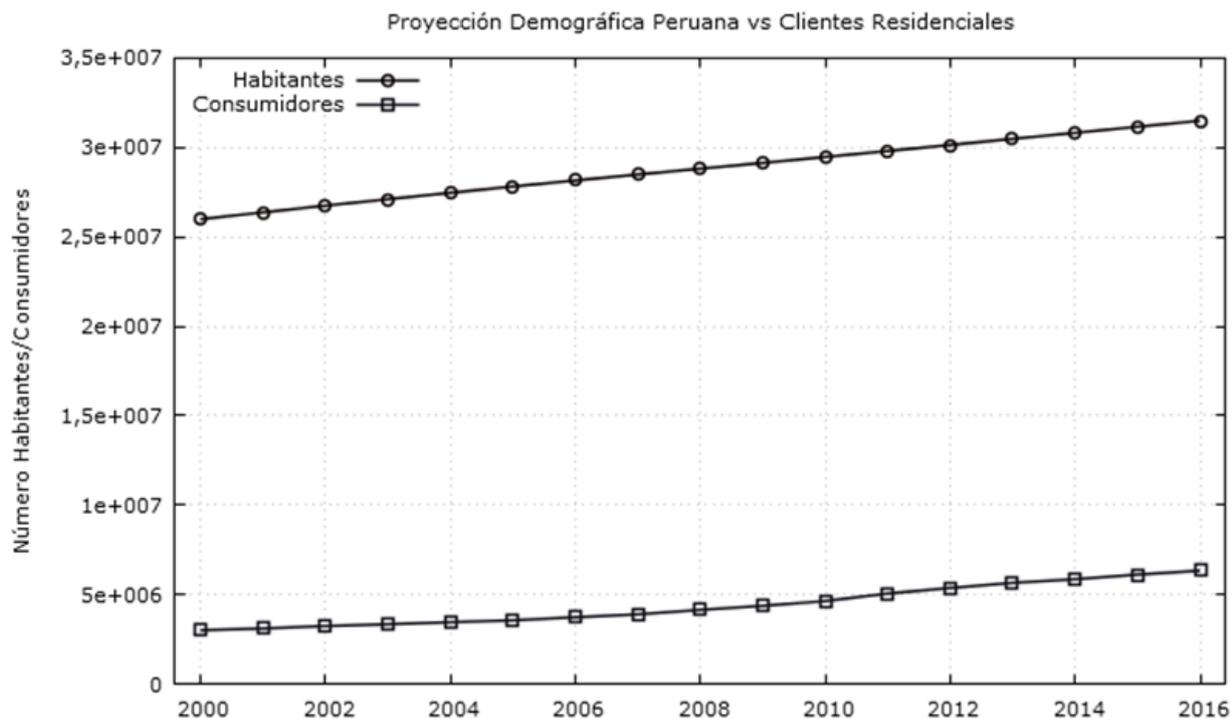


Fig. 2. Comparativo entre la proyección demográfica peruana y los clientes residenciales del sector eléctrico peruano entre los años 2000 y 2016. Fuente: Evolución de indicadores del Sector Eléctrico 1995-2016, MEM, e Instituto Nacional de Estadística e Informática INEI.

El abastecimiento a los clientes finales dentro de la infraestructura energética nacional requiere esencialmente de dos componentes importantes: Las empresas Generadoras y Distribuidoras del recurso energético. Siendo la primera de ellas de interés en el análisis de este trabajo. En la Figura 3,

observamos el número de empresas generadoras de energía eléctrica instaladas a nivel nacional desde el año 1995 al 2017, donde se puede notar claramente un incremento muy importante en el número de ellas desde el año 2015 al 2017, pasando de 150 a más de 450 empresas en solo dos años.

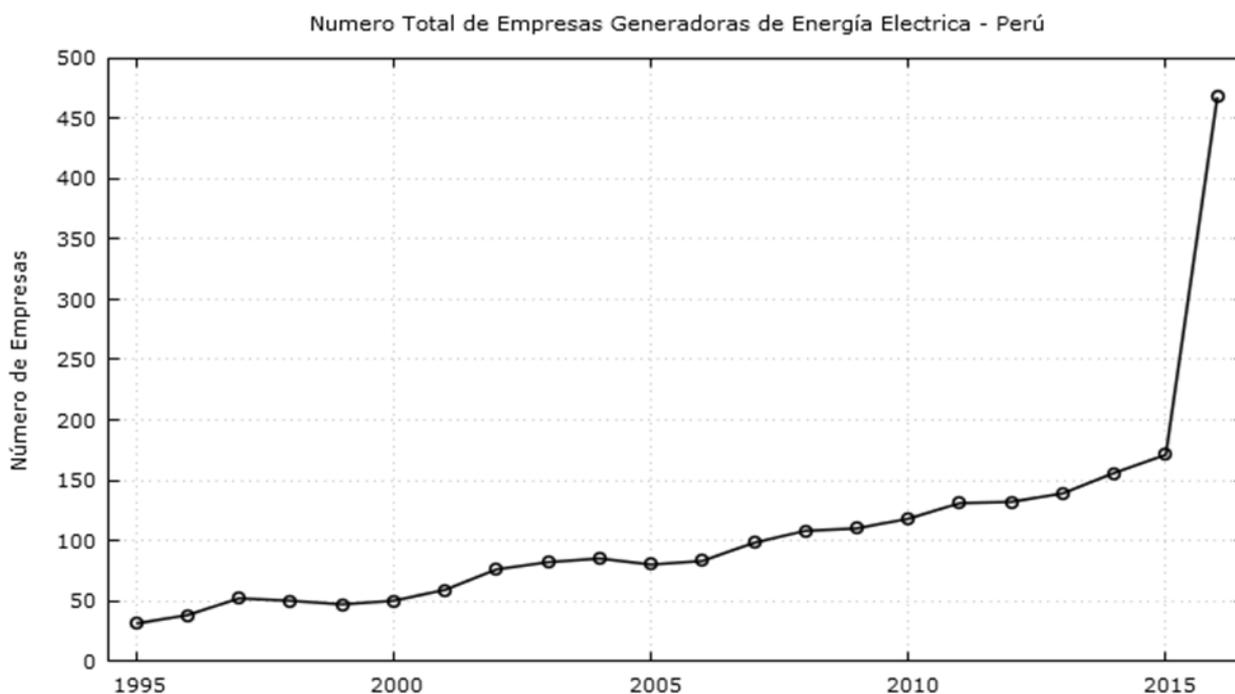


Fig. 3. Número total de empresas generadoras de energía eléctrica, correspondiente al sector eléctrico peruano. Fuente: Evolución de indicadores del Sector Eléctrico 1995-2016, MEM.

La producción de energía eléctrica de acuerdo a su fuente de origen en el sector eléctrico peruano, desde finales del siglo pasado estuvo eminentemente dominada por los sistemas hidráulicos y térmicos, quedando muy rezagados los sistemas de energías renovables: Eólica y Solar; estas últimas con participación importante en el mercado energético peruano desde el año 2012. En la Figura 4, se observar la evolución, del comportamiento y predominio de las generadoras de energía eléctrica basadas en Sistemas Hidráulicos y Térmicos, en Giga watt-hora (GW.h). Es importante mencionar

que de acuerdo a datos contenidos en el anuario de Evolución de Indicadores del Sector Eléctrico peruano, el año 2014, la producción de energía eléctrica con fuente de origen térmica comenzó a superar al de las plantas hidráulicas, es así que el año 2016 los Sistemas térmicos llegan a liderar la producción de energía eléctrica con un 50.7%, seguidos muy de cerca por los Sistemas Hidráulicos con el 46.8%, los Sistemas Eólicos con 2.1% y Sistemas Solares con un 0,5% del total de generación de energía eléctrica, como se observar en la Figura 4

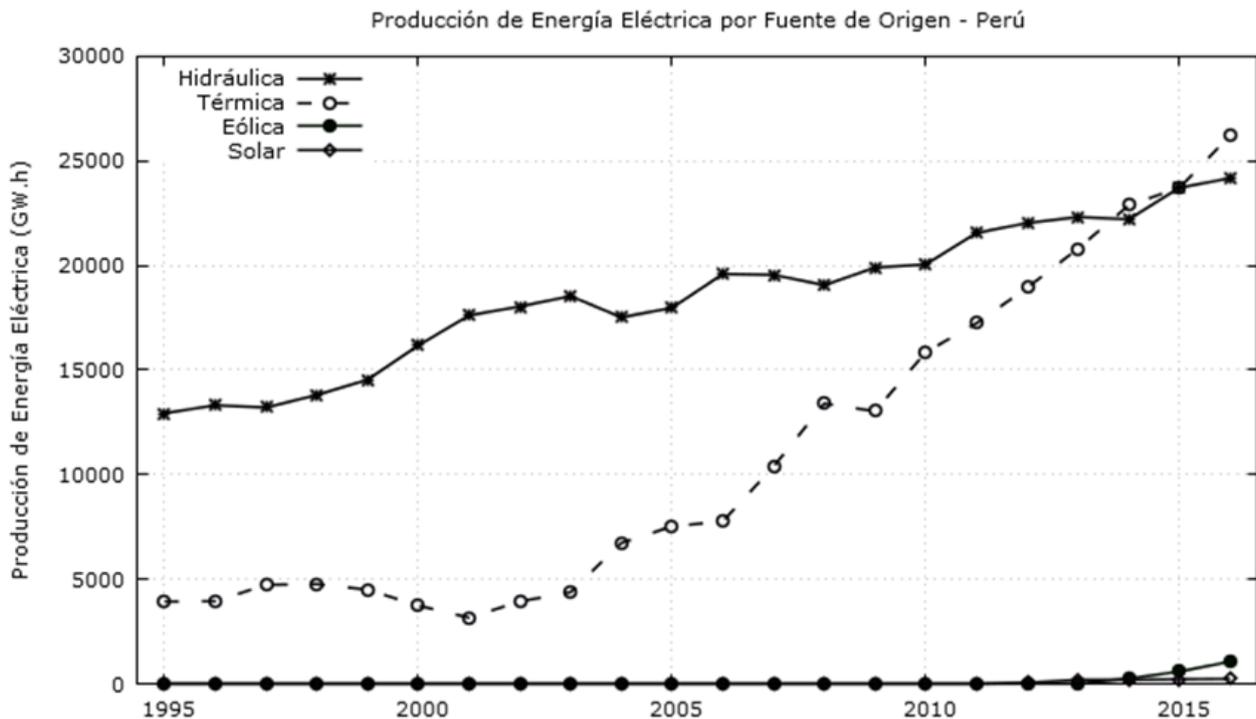


Fig. 4. Producción de Energía Eléctrica considerando la fuente de origen, correspondiente al sector eléctrico peruano. Fuente: Evolución de indicadores del Sector Eléctrico 1995-2016, MEM.

Como se había indicado inicialmente, el objetivo de este trabajo está enfocado en es establecer un modelo predictivo adecuado de la variable VGEE y que permita establecer la tendencia y comportamiento de la mencionada variable con la finalidad de establecer estrategias y medidas adecuadas para evitar el déficit energético relacionado con el desarrollo de políticas económicas y de infraestructura en el sector eléctrico peruano.

III. MODELAMIENTO: MÉTODO BOX AND JENKINS.

[1] En esta Sección, se pone de manifiesto, en esencia

la Metodología de Box and Jenkins, que como fue indicada en la Sección anterior el mismo permite ajustar el conjunto de datos paramétricos conocidos de la variable VGEE a los modelos ARIMA (Jaume A. (2001), Diseño de Series Temporales: Técnicas de Análisis. Edicions de la Universitat de Barcelona, p. 77 – 80). La aplicación de la metodología Box and Jenkins (Box, G. E. P., G. M. Jenkins,(1994), Time Series Analysis: Forecasting and Control. 3rd ed. Englewood Cliffs, NJ: Prentice Hall), establece un procedimiento el cual se puede resumir en un proceso de cinco pasos, como se puede observar en la Figura 5:

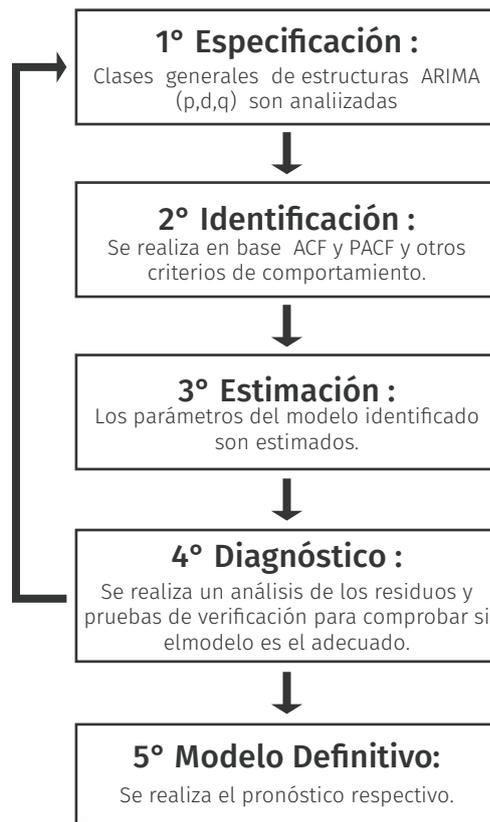


Fig. 1. Procedimiento de aplicación de la Metodología Box and Jenkins, para el modelo ARIMA (p,d,q).

Es importante indicar que en la quinta etapa del procedimiento se puede verificar la capacidad de ajuste y pronóstico del modelo. Es posible utilizar el Modelo Definitivo para realizar pronósticos basados en simulaciones utilizando el método de Montecarlo.

A. Modelos ARIMA (p, d, q).

Muchas de las series temporales encontradas cotidianamente presentan comportamientos no estacionarios, excepto aquellos relacionados a diversos tipos de “ruidos” subyacentes de algunos fenómenos físicos.

De acuerdo con lo anteriormente mencionado, en una serie temporal las diferentes observaciones y_t , se definen de forma general mediante un modelo Auto-regresivos Integrados de Medias móviles ARIMA (Autoregressive Integrated Moving average), por la siguiente expresión:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

La Metodología Box and Jenkins permite ajustar estos modelos ARIMA, a través de un ciclo iterativo, como se establece en la Figura 5. En un Modelo ARIMA (p,d,q), los parámetros p,d, y q son números enteros no negativos que indican el orden de las distintas componentes del modelo: Auto-regresiva (AR), Integrada (I) y de Medias móviles (MA). Así, una serie que modela un proceso estacionario ARMA (p, q), después de haber sido diferenciado d veces es considerada ARIMA y denotada por ARIMA (p, d, q) (Box, and Jenkins, (2016). Time

Series Analysis Forecasting and Control. Fifth Edition, Wiley Editorial, p. 192 -193). Dichos procesos a menudo se llaman procesos diferencia - estacionarios o raíz unitaria. Un modelo ARIMA expresado en su forma general tiene la forma:

$$y_t = -(\Delta^d y_t - y_t) + c + \sum_{i=1}^p \phi_i \Delta^d y_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (2)$$

Donde d representan las d-ésimas diferencias para tonar una serie en estacionaria, ϕ_1, \dots, ϕ_p , son las componentes auto-regresivos, $\theta_1, \dots, \theta_q$ son componentes de medias móviles, c es una constante, y ε_t es el término de error (denominado innovación o perturbación estocástica), conocido también como “ruido blanco del proceso”. La ecuación (2), puede ser desarrollada y expresada como:

$$\Delta^d y_t = c + \phi_1 \Delta^d y_{t-1} + \dots + \phi_p \Delta^d y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (3)$$

Introduciendo el operador de cambio retroactivo o de retrasos (desfasajes) L^i (Lag's): $L^i y_t = y_{t-i}$, en la expresión (3), obtenemos:

$$(1 - \phi_1 L^1 - \phi_2 L^2 - \dots - \phi_p L^p) \Delta^d y_t = c + (1 - \theta_1 L^1 - \theta_2 L^2 - \dots - \theta_q L^q) \varepsilon_t \quad (4)$$

Expresando los polinomios de ambos lados de la ecuación (4), en su forma compacta: $\phi(L) = (1 - \phi_1 L - \dots - \phi_p L^p)$ y $\theta(L) = (1 - \theta_1 L - \dots - \theta_q L^q)$:

$$\phi(L) \Delta^d y_t = c + \theta(L) \varepsilon_t \quad (5)$$

El operador de diferenciación se puede expresar por su equivalente $\Delta^d = \nabla^d = (1 - L)^d$, así la ecuación (5) queda como:

$$\phi(L)(1 - L)^d y_t = c + \theta(L) \varepsilon_t \quad (6)$$

En la metodología original de Box-Jenkins, es muy importante realizar la diferenciación de la serie integrada hasta que esta se torne estacionaria antes de modelar. Posteriormente, es preciso modelar la serie diferenciada como un proceso estacionario ARMA (p, q) (Box, Jenkins, and Reinsel (1994). Time Series Analysis: Forecasting and Control. 3rd ed. Englewood Cliffs, NJ: Prentice Hall). Considerando el aspecto computacional, las herramientas utilizadas en el presente trabajo son Gretl 2018b y Econometrics Toolbox de Matlab, para realizar el ajuste y el respectivo pronóstico de los procesos ARIMA (p, d, q).

B. Criterios de Información: Akaike (AIC) y Bayesiano (BIC)

Existen diferentes enfoques para establecer criterios de selección de modelos ARIMA, en este trabajo se empleará el Criterio de Información de Akaike (AIC) y el Criterio de Información de Schwarz conocido como Criterio de Información Bayesiano (BIC). Así, el criterio de información de Akaike (AIC) (Box, Jenkins, Reinsel, and Ljung (2016). Time Series Analysis Forecasting and Control. Fifth Edition, Wiley Editorial, p. 192 -193), es definido como:

$$AIC = \ln \hat{\sigma}^2 + \frac{2}{n} r \quad (7)$$

Donde, \ln es logaritmo natural, $(\hat{\sigma}^2)$ suma de cuadrados de los residuos y divididos por el número de observaciones, n

número de observaciones, r es el número total de parámetros en el modelo ARIMA.

De otro lado el criterio de información Bayesiano (BIC), es definido por:

$$BIC = \ln \hat{\sigma}^2 + \frac{\ln n}{n} r \quad (8)$$

Muy a menudo los dos criterios producen el mismo resultado, por tanto estos dos criterios serán considerados y utilizados como procedimientos adicionales en la selección del mejor modelo ARIMA.

C. Consideraciones preliminares para el Modelamiento ARIMA

Para realizar el modelamiento y respectivo análisis del comportamiento de la variable VGEE (Variable de Generación de Energía Eléctrica) utilizando el modelo ARIMA (Jaume A.G. (2001), Diseño de Series Temporales: Técnicas de Análisis.

Edicions de la Universitat de Barcelona, p. 77 – 80), es importante observar algunas características importantes de la serie a lo largo del tiempo. Estas características se encuentran las relacionadas con:

I. Tendencia y Transformación de la Serie Temporal VGEE:

En la Figura 6, se puede observar el comportamiento histórico del comportamiento de la Variable Generación de Energía Eléctrica (VGEE), que considera la producción total de energía eléctrica presentada en la Figura 4. En ella se puede observar una tendencia determinista de crecimiento exponencial sostenida a lo largo del tiempo (1995 -2016). Lo que nos sugiere realizar una transformación en la serie con la finalidad de tornarla lo más linealmente posible (alisamiento de la serie temporal), este procedimiento se realiza transformando los datos de nuestra variable VGEE utilizando la función logaritmo natural a $\ln(VGEE)$, como se puede observar en la Figura 7.

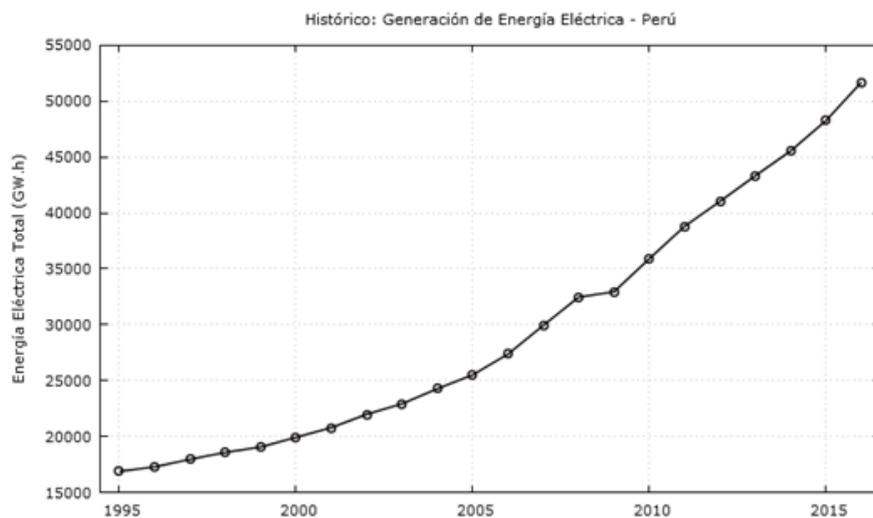


Fig. 1. Serie histórica de Generación Total de Energía Eléctrica, en el sector eléctrico peruano. Fuente: Evolución de indicadores del Sector Eléctrico 1995-2016, MEM.

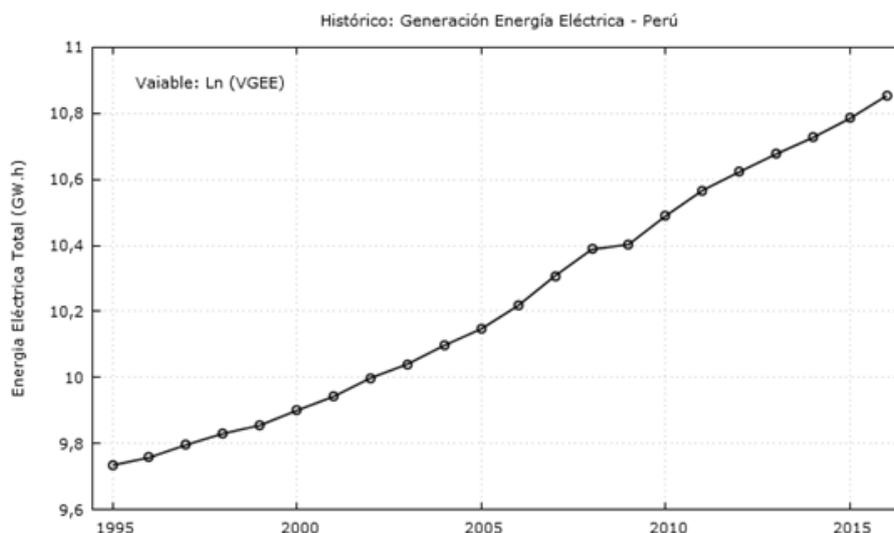


Fig. 2. Serie Histórica de la Generación de Energía Eléctrica, con transformación a partir de la función logaritmo para la variable VGEE.

II. Estadísticas Descriptivas

Considerando los datos obtenidos a partir del proceso de transformación de la variable $\ln(\text{VGEE})$, presentada en la Figura 7, algunas de las métricas estadísticas descriptivas más importantes son presentadas en la Tabla I.

TABLA I
RESUMEN DE ESTADÍSTICAS DESCRIPTIVAS PARA LA VARIABLE VGEE, EN SU FORMA LOGARÍTMICA

Media	Mediana	Mínimo	Máximo
10,233	10,182	9,7339	10,853
Desviación. Típica Estándar	C.V.	Percentil 5%	Percentil. 95%

0,36530	0,035698	9,7374	10,843
---------	----------	--------	--------

Estadística de Prueba de Normalidad a partir de la función logaritmo para la variable VGEE.

Aplicando las Funciones de Auto correlación a la variable $\ln(\text{VGEE})$, se obtuvieron valores de significancia del 1%, 5% y 10%, en las siguientes posiciones correspondientes a los retardos (lag's) para el caso de la Función de Auto-Correlación (FAC): 1, 2, 3, 4, 5, 13, 14, 15,16 y 17. En tanto que para la Función de Auto-Correlación Parcial (FACP): 1 (con significancia del 10%), como se puede ver en los correlogramas de la Figura 8. Visualmente a partir de este procedimiento es posible determinar un comportamiento decreciente en los valores de FAC, lo que nos indica claramente que la serie temporal es no estacionaria.

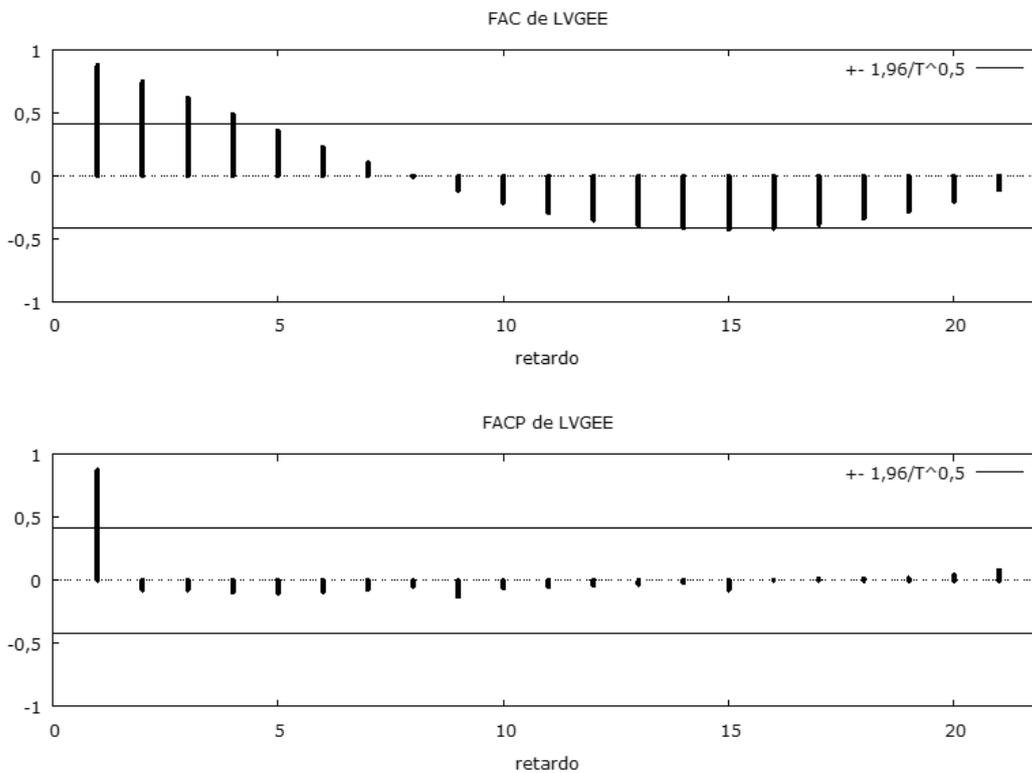


Fig. 3. Correlogramas y comportamiento de los retardos (lag's) de la FAC y FACP para la variable $\ln(\text{VGEE})$.

Con la finalidad de validar la anterior afirmación referida a que la serie de la variable $\ln(\text{VGEE})$ es no estacionaria, se aplicó la prueba o test de Hipótesis nula (H_0) de Dickey-Fuller (Morettin, Toloí (2006). *Análise de Series Temporais. 2da Edição. Editora Blucher, p.105*), desarrollada para verificar si un modelo autor regresivo posee o no raíz unitaria, que permita establecer la existencia o no la hipótesis nula (H_0), para nuestro trabajo la hipótesis nula no será rechazada para un nivel de significación mayor al 5 %, que en términos probabilidad (p), este valor estará dado por:

$p \geq 0,05$ (5%) ; (H_0) no rechazada (Serie no estacionaria)

$p < 0,05$ (5%) ; (H_0) rechazada ó (H_1) (Serie estacionaria)

Aplicando la Prueba de regresión Dickey-Fuller a la variable $\ln(\text{VGEE})$ usando el nivel respectivo de la variable (sin usar diferencias), considerando la constante y tendencia se obtuvo un valor $p= 0,3033$ que es mayor al valor de 0,05 por lo que no se rechaza (H_0), y se concluye que la serie es no estacionaria.

Volviendo a aplicar la Prueba de regresión Dickey-Fuller a la serie $\ln(\text{VGEE})$ usando la primera diferencia, considerando su constante y tendencia se obtuvo un valor $p= 0,01581$ que es un valor menor a 0,05 que rechaza (H_0), y se concluye que la serie con una diferencia consigue ser estacionaria. Por tanto, al ser aplicada la Metodología Box and Jenkins en las especificaciones de las estructuras ARIMA (p,d,q), deberán estar presentes valores de d (diferenciación) con órdenes ma-

iores a 1, es decir $d=1$ ó $d=2$.

D. Aplicación de la Metodología Box and Jenkins en estructuras ARIMA (p,d,q).

Considerando el procedimiento establecido en la Figura 5, las cinco etapas son abordadas en el presente trabajo. Para la aplicación de la Metodología propuesta, serán ejecutadas las primeras cuatro etapas (pasos del 1° - 4°, de la Figura 5), en esta primera etapa se propusieron cuatro escenarios con las siguientes estructuras: *ARIMA(1,0,1)*, *ARIMA(1,1,1)*, *ARIMA(2,1,1)* y *ARIMA(2,1,0)*. El procedimiento se realiza de forma iterativa con la finalidad de establecer la estructura más adecuada que permita realizar las predicciones de la variable (VGEE).

I. Escenario: ARIMA (1, 0,1)

En este primer escenario fue especificada una estructura ARIMA (1, 0, 1), con valores auto-regresivos (AR, $p=1$) y de medias móviles (MA, $q=1$), que teóricamente corresponden a una estructura ARMA. Se introduce este escenario con la única intención de realizar la comprobación y comparación con las otras estructuras que incluyen valores de $d \neq 0$.

La identificación del comportamiento del modelo, se realiza en base los correlogramas de FAC y FACP para los valores de retardos (Lag's), donde es posible observar que los mismos se encuentran dentro de las bandas de confianza +/- 0,5, y con valores de significancia del 1%, 5% y 10% como se observa en la Figura 9.

Respecto a los valores estadísticos obtenidos para la predicción, el Error medio es de 0,029326, la Raíz del Error cuadrático medio de 0,039772 y el Error medio porcentual de 0,28299.

Algunos de los más resaltantes valores de los parámetros estimados en base a las desviaciones típicas correspondientes al conjunto de elementos de la matriz Hessiana de modelo, permiten establecer la media para la variable Ln(VGEE) de 10,23307, media de las innovaciones 0,029326, Criterio de Información de Schwars (BIC, Criterio de Información Bayesiano) es de - 61,25580, en tanto que el Criterio de Información de Akaike (AIC, Criterio de Información de Akaike) es de - 65,61997.

II. Escenario: ARIMA (1, 1,1).

Para este segundo escenario fue especificada una estructura ARIMA (1, 1,1), para valores de auto-regresivos (AR, $p=1$), diferenciación (I, $d=1$) y de medias móviles (MA, $q=1$).

Es posible observar en los correlogramas de la FAC y FACP que los valores de retardos (Lag's) para estructura ARIMA (1, 1,1), se encuentran dentro de las bandas de confianza +/- 0,5, y con valores de significancia del 1%, y 5% como se observa en la Figura 10.

Con relación a los valores estadísticos obtenidos para la predicción, el Error medio es de 0,001768, la Raíz del Error cuadrático medio de 0,01914 y el Error medio porcentual de 0,015516.

Con relación a los parámetros más importantes estimados en base a las desviaciones típicas de los elementos de la matriz Hessiana de modelo, permiten establecer la media para la variable Ln(VGEE) de 0,053301, media de las innovaciones 0,001768, Criterio de Información de Schwars (BIC, Criterio de Información Bayesiano) es de - 94,18811, en tanto que el Criterio de Información de Akaike (AIC, Criterio de Información de Akaike) es de - 98,36620.

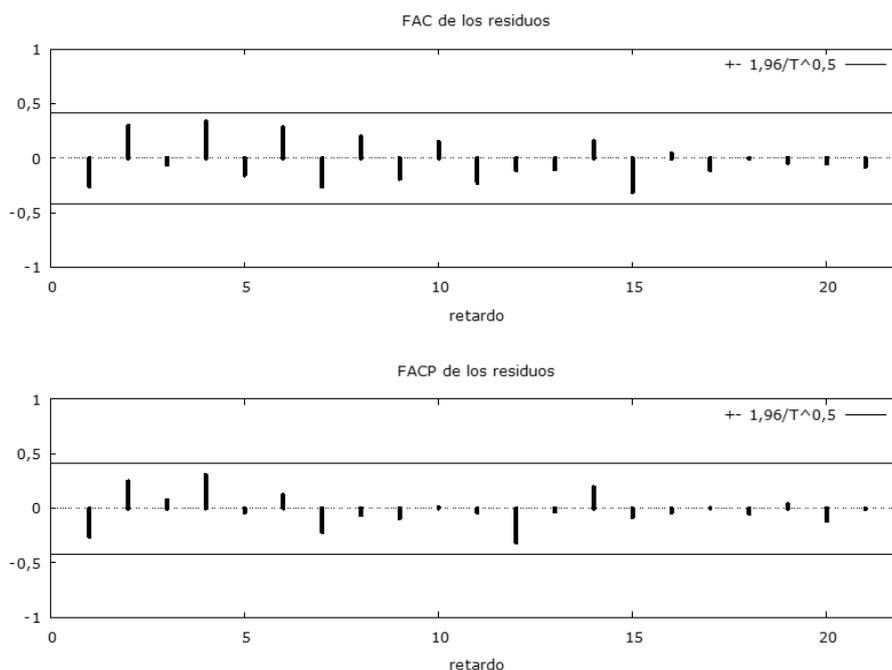


Fig. 1. Correlogramas y comportamiento de los retardos (lag's) de la FAC y FACP para la estructura ARMA (1, 0,1) para la variable Ln(VGEE).

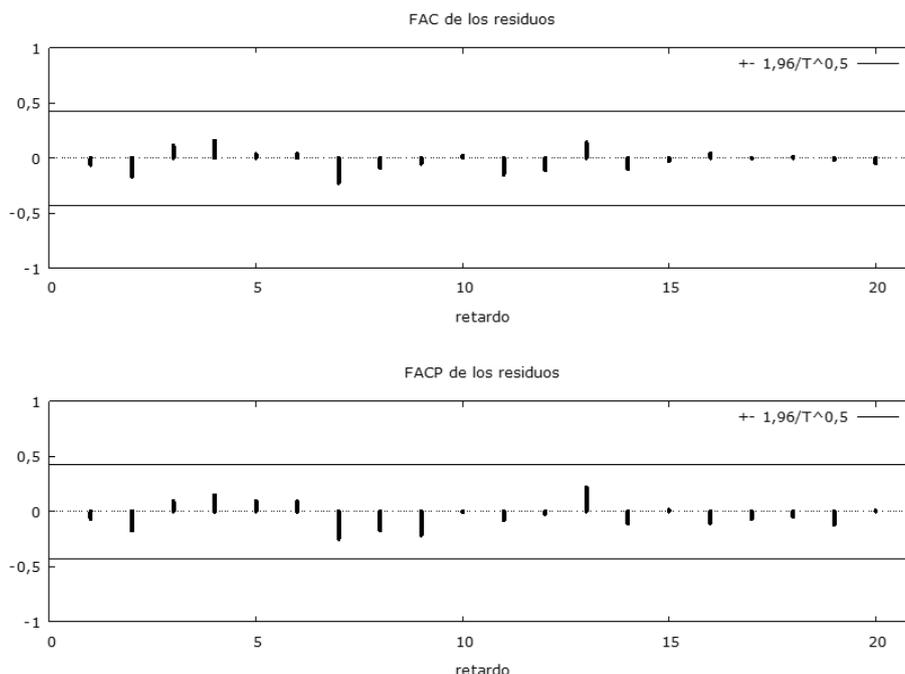


Fig. 2. Correlogramas y comportamiento de los retardos (lag's) de la FAC y FACP para la estructura ARIMA (1, 1,1) para la variable Ln(VGEE).

III. Escenario: ARIMA (2, 1,1).

En este tercer escenario fue especificada una estructura ARIMA (2, 1,1), para valores de auto-regresivos de orden 2 (AR, p=2), diferenciación (I, d=1) y de medias móviles (MA, q=1).

Considerando los correlogramas de la FAC y FACP, se puede observar que los valores de retardos (Lag's) para estructura ARIMA (2, 1,1), se encuentran dentro de las bandas de confianza +/- 0,5, y con valores de significancia del 1%, y 5% como se observa en la Figura 11.

Para esta estructura los valores estadísticos obtenidos para la predicción, corresponden al Error medio 0,0018571, la Raíz del Error cuadrático medio 0,019113 y el Error medio porcentual de 0,016448.

Considerando los parámetros más importantes estimados en base a las desviaciones típicas de los elementos de la matriz Hessiana del modelo, permiten establecer la media para la variable Ln(VGEE) de 0,053301, media de las innovaciones 0,001857, Criterio de Información de Schwars (BIC, Criterio de Información Bayesiano) es de - 91,17940, en tanto que el Criterio de Información de Akaike (AIC, Criterio de Información de Akaike) es de - 96,40201.

IV. Escenario: ARIMA (2, 1,0).

En este escenario fue especificada una estructura ARIMA (2, 1, 0), con valores auto-regresivos de orden 2 (AR, p=2), y de diferenciación (I, d=1). Observando el comportamiento de los correlogramas de la FAC y FACP para los valores de retardos (Lag's), es posible observar también como en los escenarios anteriores que los valores obtenidos se encuentran dentro de las bandas de confianza +/- 0,5, y con valores de significancia del 1%, y 5% como se observa en la Figura 12.

Los valores estadísticos obtenidos para la predicción, corresponden al Error medio 0,00044497, la Raíz del Error cuadrático medio 0,019596 y el Error medio porcentual de 0,0016881.

Algunos de los parámetros más importantes estimados en base a las desviaciones típicas de los elementos de la matriz Hessiana del modelo, permiten establecer la media para la variable Ln(VGEE) de 0,053301, media de las innovaciones 0,000445, Criterio de Información de Schwars (BIC, Criterio de Información Bayesiano) es de - 93,33487, en tanto que el Criterio de Información de Akaike (AIC, Criterio de Información de Akaike) es de - 97,51296.

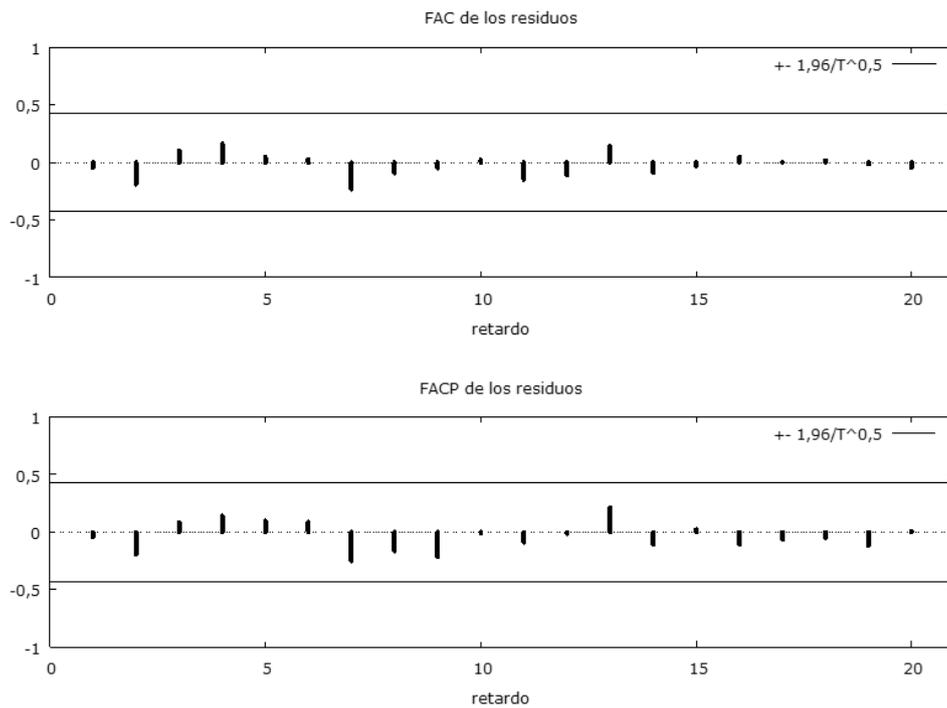


Fig. 3. Correlogramas y comportamiento de los retardos (lag's) de la FAC y FACP para la estructura ARIMA (2, 1, 1) para la variable Ln(VGEE).

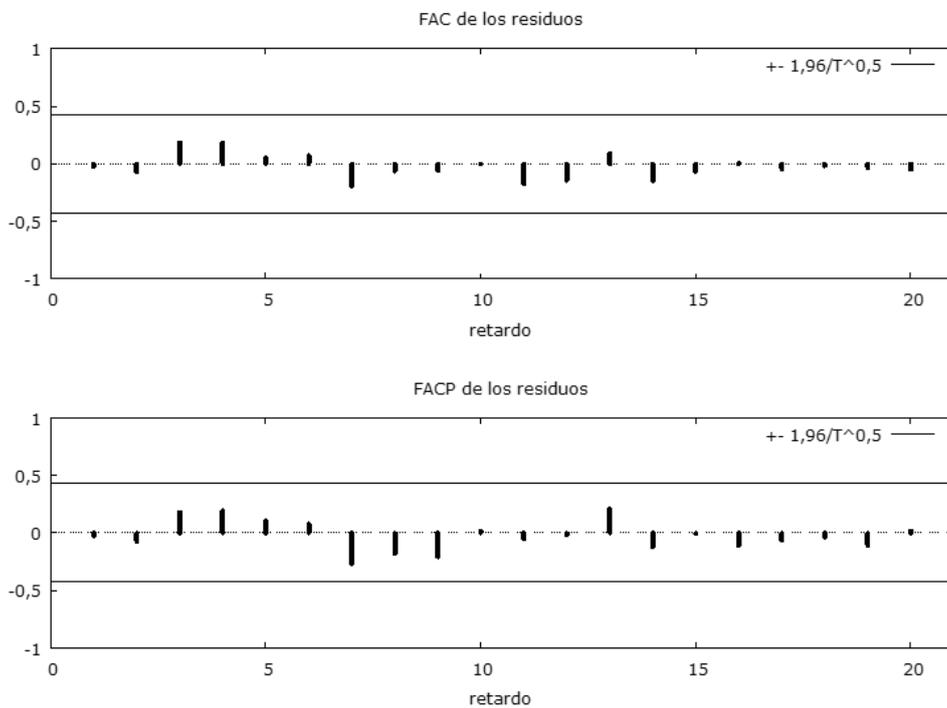


Fig. 4. Correlogramas y comportamiento de los retardos (lag's) de la FAC y FACP para la estructura ARIMA (2, 1, 0) para la variable Ln(VGEE).

IV. ANÁLISIS Y RESULTADOS NUMÉRICOS.

Considerando la primera etapa correspondiente a la Metodología establecida en la Figura 5, para los cuatro escenarios evaluados en el ítem C de la III parte de este trabajo, en la Tabla 2 se presentan en resumen los parámetros más importan-

tes estimados y diagnosticados a partir de los resultados obtenidos de la matriz Hessiana, para cada estructura ARIMA.

TABLA 2
RESUMEN DE PARÁMETROS CORRESPONDIENTES
A LOS CUATRO ESCENARIOS ARIMA

	ARIMA (1, 0, 1)	ARIMA (1, 1, 1)	ARIMA (2, 1, 1)	ARIMA (2, 1, 0)
Media de la Variable	10,23307	0,053301	0,053301	0,053301
Media de Innovaciones	0,029326	0,001768	0,001857	0,000445
Criterio de Schwarz (BIC)	-61,25580	-94,18811	-91,17940	-93,33487
Criterio de Akaike (AIC)	-65,61997	-98,36620	-96,40201	-97,51296
Error Medio	0,029326	0,001768	0,0018571	0,00044497
Raíz del Error Cuadrático Medio	0,039772	0,01914	0,019113	0,019596
Error Medio Porcentual	0,28299	0,015516	0,016448	0,0016881

A. Consideraciones preliminares para el Análisis de la Metodología Box and Jenkins en estructuras ARIMA (p,d,q).

Algunas de las consideraciones establecidas por John E. Hanke y Dean Wichern (*Morettin, Toloí (2006). Análise de Series Temporais. 2da Edição. Editora Blucher, 2006*), para la selección de los Modelos (estructuras) ARIMA, se han desarrollado varios enfoques relacionados con el número de parámetros que estos utilizan. Pero son también de mucha importancia el uso de Criterios de Información con la finalidad de seleccionar dichos Modelos ARIMA, en este trabajo son considerados los Criterios de Información de Akaike (AIC) y Schwarz (BIC), basados en penalizaciones y sugiere que el uso de valores pequeños en estos criterios para realizar la selección del Modelo adecuado. Para el caso que los valores de AIC y BIC correspondan a valores negativos en este caso se consideran los valores negativos más altos como el Modelos adecuado (*Emiliano P.C.(2009) Fundamentos e aplicações dos critérios de informação: Akaike e Bayesiano (Dissertação). Universidade Federal de Lavras. Minas Gearis-Brasil*). Otros parámetros de mucha utilidad son los relacionados con valores de Error Medio, Raíz del Error Cuadrático Medio y el Erros Medio Porcentual, con valores muy bajos, próximos a cero.

B. Análisis y Resultados.

De acuerdo a las consideraciones indicadas en el párrafo anterior, y los parámetros obtenidos en la Tabla 2, podemos observar que los Modelos ARIMA (1, 1, 1) y ARIMA (2, 1, 0) son aquellos con mejor desempeño y cumplen con los criterios establecidos para este trabajo. Se observa también que para el caso del Modelo ARIMA (2, 1, 0) los valores de Medios de Innovaciones (errores) y el Error Medio son bastante pequeños con respecto al Modelo ARIMA (1, 1, 1), que no

necesariamente garantizan que el Modelo ARIMA (2, 1,0) sea el más óptimo. De otro lado, un aspecto importante en la elección del tipo de modelo está fuertemente relacionada con los Criterios de Información obtenidos en AIC y BIC, y observamos que los valores más negativos corresponden al Modelo ARIMA (1, 1, 1). Considerando la diferencia de los valores de los parámetros entre ambos modelos, estos valores son presentados en la Tabla 3.

TABLA 3
RESUMEN DE DIFERENCIAS DE VALORES DEL
MODELOS ARIMA (1, 1, 1) Y ARIMA (2, 1, 0)

	ARIMA (1, 1, 1)	ARIMA (2, 1, 0)	Diferencia
Media de la Variable	0,053301	0,053301	0,000000
Media de Innovaciones	0,001768	0,000445	0,001323
Criterio de Schwarz (BIC)	-94,18811	-93,33487	-0,85324
Criterio de Akaike (AIC)	-98,36620	-97,51296	-0,85324
Error Medio	0,001768	0,00044497	0,00132303
Raíz del Error Cuadrático Medio	0,01914	0,019596	-0,000456
Error Medio Porcentual	0,015516	0,0016881	0,0138279

Finalmente, basados en los valores de diferencia de ARIMA(1,1,1) y ARIMA(2,1,0) de la tercera columna en la Tabla 3, optamos por elegir el Modelo ARIMA(2,1,0) como Modelo Definitivo para realizar la predicción de la Variable Ln(VGEE), ya que esta última presenta un buen desempeño en la predicción con valores bajos de Errores Medio, Raíz del Error Cuadrático Medio y Error Medio Porcentual, y una diferencia muy pequeña del orden de - 0,85324 que de forma práctica se pueden considerar equivalentes para los valores de los criterios de información AIC y BIC. Los valores de los parámetros del Modelo ARIMA(2,1,0): $c=0,0530414$, $\phi_1= 0,205951$, $\phi_2= 0,0489914$, y el valor medio de $\varepsilon_t = 0,000445$.

Expandiendo la ecuación compacta (6) para el Modelo ARIMA (2, 1,0):

$$(1 - \phi_1 L^1 - \phi_2 L^2)(1 - L)^d y_t = c + (1) \varepsilon_t \quad (9)$$

Reemplazando los valores de los parámetros obtenidos en (7), se obtiene la ecuación del Modelo Definitivo:

$$(1 - 0,205951L^1 - 0,0489914L^2)(1 - L)^1 y_t = 0,0530414 + 0,000445 \quad (10)$$

En la Tabla 4, se observan los valores de la Variable Ln(-VGEE), las estimativas y los respectivos residuos correspondientes a la serie desde el año 1996 – 2017, observando que el modelo ARIMA (2,1,0) reproduce de forma adecuada el comportamiento histórico de la serie de la variable Ln(V-GEE). En la Figura 13

TABLA 4
ESTIMATIVAS Y RESIDUOS DE LA VARIABLE
ANALIZADA CON EL MODELO ARIMA (2, 1, 0)

Año	Ln(VGEE)	Valor Estimado	Residuos
1996	9,75729	9,78619	-0,0289
1997	9,79554	9,80391	-0,0084
1998	9,82998	9,84408	-0,0141
1999	9,8548	9,87846	-0,0237
2000	9,89961	9,90112	-0,0015
2001	9,94202	9,94958	-0,0076
2002	9,99799	9,99247	0,00552
2003	10,0399	10,0511	-0,0112
2004	10,0969	10,0908	0,00607
2005	10,1468	10,1502	-0,0034
2006	10,2172	10,1994	0,01779
2007	10,3071	10,2737	0,0334
2008	10,3879	10,3685	0,01933
2009	10,4026	10,4484	-0,0458
2010	10,4887	10,4491	0,03962
2011	10,5663	10,5467	0,01965
2012	10,6222	10,6261	-0,0039
2013	10,6766	10,677	-0,0004
2014	10,7266	10,7301	-0,0035
2015	10,7846	10,779	0,00554
2016	10,8532	10,8385	0,01472

Para realizar la predicción de la variable Ln(VGEE) correspondiente al Recurso Generación de energía eléctrica, en el Sector Eléctrico peruano, consideramos una predicción conservadora con un horizonte de 5 años a partir del año 2017 al 2021. Es importante indicar que los datos oficiales de generación de energía eléctrica del año 2017 y lo que va del año 2018, no se encuentran publicados oficialmente.

En la Tabla 5, se puede apreciar la predicción de la variable Ln(VGEE), para un horizonte de cinco años y los respectivos parámetros como la desviación típica y el intervalo de confianza del 95%.

TABLA 5
PREDICCIÓN CON HORIZONTE DE 5 AÑOS PARA LA
VARIABLE ANALIZADA CON EL MODELO ARIMA (2, 1, 0)

Observaciones	Ln(VGEE)	Predicción	Desviación Típica	Intervalo de 95%
2017		10,9097	0,0195964	(10,8713, 10,9481)
2018		10,9642	0,0307002	(10,9041, 11,0244)
2019		11,0177	0,0398605	(10,9396, 11,0959)
2020		11,071	0,0475852	(10,9777, 11,1642)
2021		11,1241	0,0543186	(11,0176, 11,2305)

En la Figura 13, son presentados los valores obtenidos a partir del Modelo ARIMA (2, 1, 0), y especificadas en la Tabla 4 y Tabla 5.

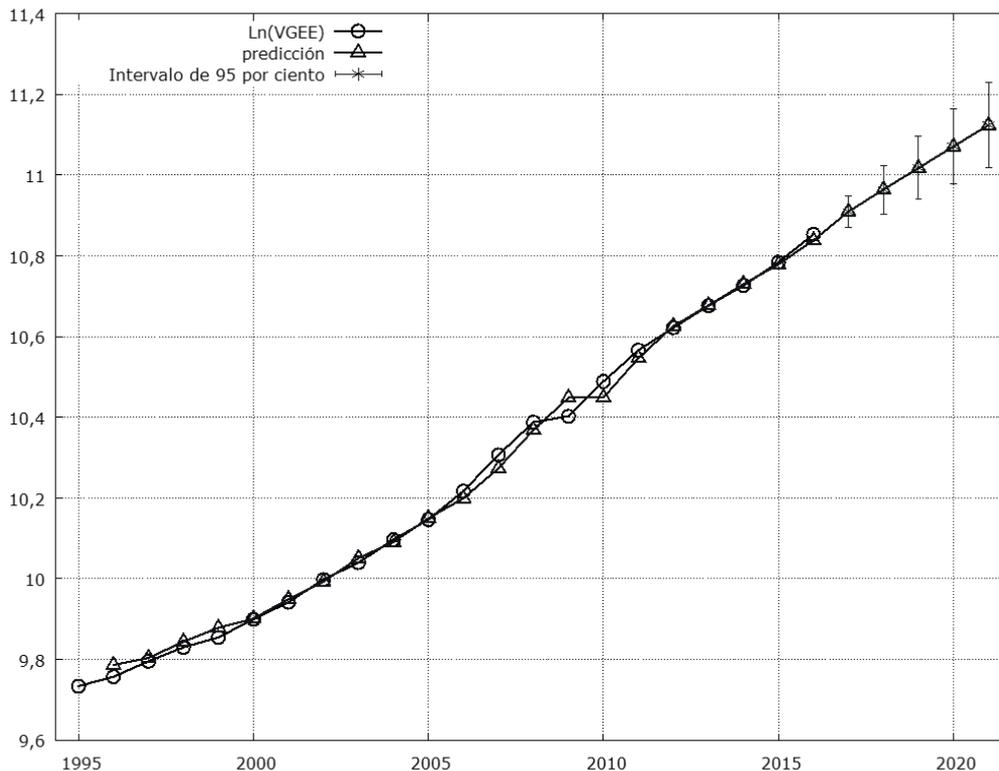


Fig. 1. Predicción, Estimativas e intervalo de confianza al 95% de la Variable Ln(VGEE), para un horizonte predictivo de 5 años.

Los datos presentados en la Tabla 5 corresponden a valores de la Variable VGEE expresadas como logaritmo natural, es decir $\ln(VGEE) = \ln(GW.h)$. Los valores Transformados a $GW.h$, son presentados en la Tabla 6.

TABLA 6
 PREDICCIÓN CON HORIZONTE DE 5
 AÑOS PARA LA VARIABLE VGEE

Observaciones	Predicción VGEE (GW.h)	Desviación Típica	Intervalo de 95%
2017	54704,43	1,0197	(52643,602, 56845,935)
2018	57768,56	1,0311	(54398,942, 61353,039)
2019	60943,34	1,0406	(56364,792, 65900,413)
2020	64279,75	1,0487	(58553,725, 70558,681)
2021	67785,25	1.0558	(60937,254, 75395,283)

V. CONCLUSIONES.

Se pudo observar un buen desempeño de los Modelos ARIMA (1, 1, 1) y ARIMA (2, 1, 0); y considerando los pronósticos respectivamente se pudo observar que los valores de error del Modelo ARIMA (2, 1, 0) fueron menores que los obtenidos con el Modelo ARIMA (1, 1, 1), que permitieron establecer como Modelo Definitivo para la predicción de la variable VGEE (GW.h).

Con respecto a los valores obtenidos en la predicción de la variable Generación de energía eléctrica (VGEE), presentados en la Tabla 6, es importante observar que, los incrementos porcentuales de generación de energía eléctrica para el sector eléctrico peruano a partir del último dato oficial y de dominio público correspondiente al año 2016 (51699,97 GW.h) serán los siguientes: al 2017 se deberá incrementar la generación eléctrica en 5.81% , al año 2018 se deberá incrementar en 11.73%, para el año 2019 el incremento deberá de bordear el 17.87%, para el año 2020 este recurso deberá alcanzar un incremento de 24.33% y para el año 2021 el mismo deberá encontrarse en torno de 31.11% , con la finalidad de asegurar el suministro energético en condiciones adecuadas para el país.

Considerando que la Producción de Energía Eléctrica (GW.h) al año 2016 mostrada de la Figura 4, y expresada en términos porcentuales, es posible deducir que, las Generadoras de energía eléctrica Térmica poseen el 50.7% de Producción de la Energía Eléctrica en el país, seguida de los Sistemas Hidráulicos el 46.8%, Sistemas Eólicos 2.1% y Sistemas Solares con un 0,5%. De otro lado, con base en el análisis y los resultados obtenidos en el presente trabajo se espera que para el año 2021 se tenga un crecimiento sostenido y proporcional por encima del 30% en cada uno de los sistemas, refle-

jándose en cada uno de ellos una expansión en su producción y capacidad operativa. También este escenario permite establecer un fuerte incremento en la inversión económica, de infraestructura eléctrica y logística por parte del estado y la empresa privada, permitiendo la creación de nuevos puestos laborales en el ámbito del sector eléctrico peruano, en los siguientes años venideros coincidiendo con la conmemoración del bicentenario de la independencia del Perú.

BIBLIOGRAFÍA

Patricia Vélez, “Mineras de Perú, en alerta energética”, Redacción de Reuters. (<https://lta.reuters.com/article/top-News/idLTASIE81K03M20120221?sp=true>), 21 de febrero de 2012.

Mariana Mazo, “Vizcarra asegura que impondrá proyectos mineros a comunidades que no lo deseen”, Europapress/Internacional (<http://www.europapress.es/internacional/noticia-vizcarra-asegura-no-impondra-proyectos-mineros-comunidades-no-deseen-20180419015100.html>), Lima 19 de abril del 2018.

Ministerio de Energía y Minas- Perú. Evolución de Indicadores del Sector Eléctrico 1995-2016. Dirección General de Electricidad, Dirección de Estudios y Promoción Eléctrica, 2017.

Jaume Arnau i Gras. Diseño de Series Temporales: Técnicas de Análisis. Edicions de la Universitat de Barcelona, Primera Edición 2001, p. 77 – 80.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. Time Series Analysis: Forecasting and Control. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1994.

George E.P. Box, Gwilym M Jenkins, Gregory C. Reinsel, Greta M. Ljung. Time Series Analysis Forecasting and Control. Fifth Edition, Wiley Editorial, 2016,p. 192 -193.

Pedro A. Morettin, Cléia M. C. Toloí. Análise de Series Temporais. 2da Edição. Editora Blucher, 2006, 105, p.

John E. Hanke, Dean Wichern. Business Forecasting. 9va Edition. Pearson New International Edition, 2014, 387,p.

Dickey, D.A. and W.A. Fuller (1979), “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”, Journal of the American Statistical Association, 74, p. 427–431

Paulo César Emiliano. Fundamentos e aplicações dos critérios de informação: Akaike e Bayesiano (Disertação). Universidade Federal de Lavras. Minas Gearis-Brasil, 2009.