

Análisis exploratorio entre modelos matemáticos predictivos, aplicados a la producción de energía mediante series temporales

Exploratory analysis between predictive mathematical models, applied to energy production through time series

- ¹ Guido Javier Mazón Fierro  <https://orcid.org/0000-0001-8745-2373>
Escuela Superior Politécnica de Chimborazo (ESPOCH), Facultad de Administración de Empresas, Riobamba, Ecuador,
guido.mazon@epoch.edu.ec
- ² Pamela Alexandra Buñay Guisñan  <https://orcid.org/0000-0002-4320-6899>
Universidad Nacional de Chimborazo (ESPOCH), Facultad de Ingeniería, Riobamba, Ecuador
pbunay@unach.edu.ec



Artículo de Investigación Científica y Tecnológica

Enviado: 11/04/2022

Revisado: 20/05/2022

Aceptado: 27/06/2022

Publicado: 13/07/2022

DOI: <https://doi.org/10.33262/concienciadigital.v5i3.1.2223>

Cítese:

Mazón Fierro, G. J., & Buñay Guisñan, P. A. (2022). Análisis exploratorio entre modelos matemáticos predictivos, aplicados a la producción de energía mediante series temporales. *ConcienciaDigital*, 5(3.1), 57-78.
<https://doi.org/10.33262/concienciadigital.v5i3.1.2223>



CONCIENCIA DIGITAL, es una revista multidisciplinar, **trimestral**, que se publicará en soporte electrónico tiene como **misión** contribuir a la formación de profesionales competentes con visión humanística y crítica que sean capaces de exponer sus resultados investigativos y científicos en la misma medida que se promuevan mediante su intervención cambios positivos en la sociedad. <https://concienciadigital.org>

La revista es editada por la Editorial Ciencia Digital (Editorial de prestigio registrada en la Cámara Ecuatoriana de Libro con No de Afiliación 663) www.celibro.org.ec

Esta revista está protegida bajo una licencia Creative Commons Attribution Non Commercial No Derivatives 4.0 International. Copia de la licencia: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Palabras claves:

Series temporales, pronóstico, energía, modelos matemáticos.

Resumen

Introducción: la energía en los actuales momentos se puede considerar como un elemento esencial en la vida de las personas, así como en el desarrollo y progreso de los países, el sector energético se constituye como estratégico debido a que permite el funcionamiento y operabilidad de los diferentes sectores se puede afirmar que la energía es indispensable en la sociedad moderna. El pronosticar o inferir que va a suceder a futuro, permite tomar decisiones oportunas y anticiparse a los acontecimientos, es así como se vuelve trascendente el conocer la producción del sector energético a futuro, además, se pueden utilizar estas predicciones como elementos de partida para generar documentos como planificaciones energéticas a mediano y largo plazo. **Objetivos:** realizar un estudio exploratorio de las mejores técnicas que podrían asistir la predicción en la producción de energía primaria en Ecuador, para evaluar la eficiencia de ajuste a corto plazo mediante series temporales univariantes. **Metodología:** en el trabajo investigativo se pudo realizar un estudio exploratorio de cuatro modelos predictivos en el sector energético de Ecuador, mediante dos técnicas, ARIMA y suavización exponencial Holt, que permitieron una aproximación confiable de predicción en la producción de energía primaria a corto plazo, en tres años hasta 2022, mediante series temporales univariantes. En cuanto a la parte metodológica empleada para cumplir los objetivos, se inició con la obtención de la serie histórica proporcionada por el Ministerio de Recursos Renovables y Energía en el documento técnico denominado Balance Energético Nacional 2019, se procesaron los datos y determinaron outliers mediante el criterio de Chauvenet, una vez determinada la base de datos para el análisis, se aplicó la metodología Box-Jenkins para la obtención de modelos ARIMA y Holt. **Resultados:** el modelo que mejor se ajusta a las bondades de predicción de los analizados es el Modelo_a ARIMA (1,1,0) cuya expresión es: $Y_t = 3365,526 + 0,074 Y_{t-1} + \varepsilon_t$, además, se estimó que la producción de energía primaria para el año 2022 en Ecuador, podría ser de 236940,541 kilo barriles equivalentes de petróleo (KBEP), con una fluctuación superior e inferior en el intervalo de [27511,589 , 198369,493](KBEP). **Conclusiones:** se puede afirmar de acuerdo con los datos obtenidos que los modelos predictivos hallados son estrictamente autorregresivos es decir que son métodos iterativos explícitos, puesto que determinan

el valor de Y_t en dependencia con el anterior resultado Y_{t-1} , en el cual no intervienen los residuos de los errores, esto indica que no interviene la componente de medias móviles. La predicción con los tres primeros modelos a, b, c resultaron con un comportamiento creciente y con el modelo h se mantenía constante.

Keywords:

Typography:
Time series,
forecast, energy,
mathematical
models.

Abstract

Introduction: energy at the present time can be considered as an essential element in people's lives, as well as in the development and progress of countries, the energy sector is strategic because it allows the operation and operability of the different sectors. it can be said that energy is indispensable in modern society. Forecasting or inferring what is going to happen in the future allows timely decisions to be made and to anticipate events. This is how it becomes important to know the production of the energy sector in the future. In addition, these predictions can be used as starting elements to generate documents such as medium and long-term energy planning. **Objectives:** conduct an exploratory study of the best techniques that could assist in the prediction of primary energy production in Ecuador, to evaluate the short-term adjustment efficiency through univariate time series. **Methodology:** in the research work it was possible to conduct an exploratory study of four predictive models in the energy sector of Ecuador, using two techniques, ARIMA and Holt exponential smoothing, which allowed a reliable approximation of prediction in the production of primary energy in the short term, in three years until 2022, using univariate time series. As for the methodological part used to meet the objectives, it began with obtaining the historical series provided by the Ministry of Renewable Resources and Energy in the technical document called National Energy Balance 2019, the data was processed and outliers were determined using the criterion de Chauvenet, once the database for the analysis was determined, the Box-Jenkins methodology was applied to obtain ARIMA and Holt models. **Results:** the model that best fits the prediction benefits of those analyzed is ARIMA Model-a (1,1,0) whose expression is: $Y_t = 3365.526 + 0.074 Y_{(t-1)} + \varepsilon_t$, in addition, it was estimated that Primary energy production for the year 2022 in Ecuador could be 236940.541 kilobarrels of oil equivalent (KBEP), with a fluctuation above and below in the interval of [275511.589 .198369.493](KBEP). **Conclusions:** based on the data obtained, it

can be stated that the predictive models found are strictly autoregressive, that is, they are explicit iterative methods, since they determine the value of Y_t depending on the previous result $Y_{(t-1)}$, in which they do not intervene the residuals of the errors, this indicates that the component of moving averages does not intervene. The prediction with the first three models a, b, c resulted in an increasing behavior and with model h it remained constant.

Introducción

La energía en los actuales momentos se puede considerar como un elemento esencial en la vida de las personas y en el desarrollo y progreso de los países, el sector energético se constituye como estratégico debido a que permite el funcionamiento y operabilidad de los diferentes sectores, como por ejemplo industrial, transporte, agrícola, residencial, entre otros, entonces se puede decir que la energía está presente y se manifiesta en la cotidianidad de la vida de los seres humanos cuando en el ámbito productivo se opera una maquinaria, o en una residencia se utiliza energía eléctrica para iluminar una habitación o encender un electrodoméstico, también se puede evidenciar la presencia de energía cuando se desea transportar de un lugar a otro mediante la utilización de un medio de transporte motorizado en síntesis se puede afirmar que la energía es indispensable en la sociedad moderna.

El pronosticar o inferir que va a suceder a futuro, permite tomar decisiones oportunas en el momento adecuado y anticiparse a los acontecimientos, es por esto por lo que se vuelve trascendente el conocer la producción y el consumo del sector energético, además de esto se puede utilizar estas predicciones como elementos de partida para generar documentos más elaborados como una planificación energética a mediano y largo plazo. Si se asocia la predicción a una estructura matemática se puede afirmar según González (2009), que el modelo predictivo es una representación de una realidad compleja, que se diseña para analizar su comportamiento y anunciar o conjeturar lo que en determinadas condiciones ha de suceder. El insumo con el que se debe contar para establecer proyecciones en un estudio son datos del pasado es decir una serie temporal, la cual es el resultado de observar los valores de una variable a lo largo del tiempo en intervalos regulares cada día, mes, año (Alonso, 2019), si a esto le añadimos el análisis univariante o escalar según Peña (2010), es una secuencia de N datos ordenados y equidistantes cronológicamente sobre una característica de una unidad observable en diferentes momentos.

El poder desarrollar un caso de estudio particular con datos históricos de producción de energía primaria en Ecuador, proporcionados por el Ministerio de Recursos Renovables y Energía, brindan un aporte al sector energético, ya que contar con valores o estimaciones futuras basados en evidencia científica siguiendo un modelo predictivo ARIMA y Holt, son insumos importantes y necesarios para realizar una planificación adecuada en el sector, además de ser un importante indicador en la toma de decisiones. Estas acciones permitirán tener un abastecimiento apropiado y no permitirán llegar a producir un déficit de energía, que podrían afectar a otros sectores del país, además de reducir el coste de importación con relación a la compra anticipada del recurso.

El problema que se pretende ayudar a solventar es estimar la producción de energía primaria en un periodo corto de tres años para poder obtener un valor aproximado de la producción en el año 2022.

Metodología

Recopilación y preparación de datos

El punto de partida se fundamenta en la recopilación de los datos históricos de la producción de energías primarias en Ecuador, la fuente de donde se tomaron los mencionados datos es el Balance Nacional de Energía (BNE), el cual es difundido por el Ministerio de Energía y Recursos Naturales no Renovables en conjunto con otras organizaciones del sector energético de Ecuador, cabe mencionar que la información es de libre acceso la cual se encuentra en el portal web de la institución. Esta serie de datos cronológicos se los presenta de dos maneras en el Balance Nacional de Energía, la primera mediante gráficas que comprende un periodo anual de 1970 hasta 2012, y la segunda a través de tablas de valores desde 1995 hasta el año 2019, lo que comprende una muestra de 50 datos para el análisis.

Para poder cuantificar la serie temporal de datos que se muestran de 1970 al 2012 en la gráfica del Balance Nacional de Energía, se ha utilizado el programa de libre acceso webplotdigitizer. El procedimiento que se desarrolló para la digitalización de los datos fue:

- Obtener la gráfica a digitalizar en formato *.jpg para poder importar al menú principal de webplotdigitizer.
- Definir el punto de origen de la gráfica, lo que equivale a establecer la coordenada $x, y (0,0)$
- Determinar el valor máximo para cada eje coordenado en $x = 2012 ; y = 250\ 000$

- Una vez determinados estos pasos el programa establece un mallado en el cual se asignará un valor numérico cuando se posicione el puntero del ratón utilizando puntos donde se desee conseguir la coordenada (x, y)
- Se procede a sobre escribir la gráfica con puntos consecutivos para sacar los valores deseados.
- Para finalizar se debe exportar el archivo que contiene la información de la gráfica, transformada a una tabla de datos numéricos.

Una vez agrupados todos los valores de producción de energía primaria en una serie temporal anual de 1970 hasta el 2019 se realiza una gráfica de líneas para visualizar cómo se comporta la serie de tiempo.

En la preparación de los datos se incluyó el análisis de la serie temporal, en un primer momento elaborando un diagrama de cajas y bigotes para identificar los outliers, además de esta ayuda gráfica se aplicó el criterio de Chauvenet.

Determinados los candidatos a outliers se procede a la elección mediante el criterio de Chauvenet el cual consiste en:

- Se debe transformar una serie de datos con tendencia a una serie temporal estacionaria.
- Como siguiente paso se calcula los estadísticos descriptivos de la serie de datos estacionaria.
- Se estima el punto crítico de Chauvenet calculado (P_i) para todos los valores de la serie temporal.
- A continuación, se compara cada valor de P_i con el punto crítico de Chauvenet teórico (P_{ch}), si $P_i > P_{ch} \Rightarrow$ es un outliers o valor atípico

Una vez identificados los valores atípicos se procede a entender el origen y causas de su presencia y a tomar la decisión de eliminarlo, modificarlo o trabajar con ese dato fuera de rango.

Desarrollo y formulación de los modelos predictivos

Con los valores de la data explicados, en esta etapa se procede primero a graficarlos, para en primera instancia observar características propias de esta serie temporal univariante, segundo con la ayuda del software SPSS se procede al análisis de características de la serie como la función de auto correlación simple y parcial además de identificar si la serie es estacionaria o no, con esta información obtenida se procede a estructurar modelos lineales tales como ARIMA con diferentes variantes, por otro lado con los mismos datos se aplicara modelos con suavizamiento exponencial o de Holt. El acrónimo ARIMA tiene su origen en el término anglosajón *Autoregressive Integrated Moving Average* que

significa Auto Regresivo (AR) Integrado (I) Medias Móviles(MA), llamado también como modelo Autorregresivo de medias móviles integrado; la publicación de George Box y Gwilym Jenkins , marcó el comienzo de una nueva generación de herramientas de pronóstico la misma que dio origen a la metodología de Box-Jenkins (BJ), pero que también es conocida como metodología ARIMA (Herrera, 2019), una de las ventajas que se puede citar es que son muy buenos para la predicción a corto plazo debido a su capacidad de aprender de los cambios de la serie, además de lo mencionado se puede decir que la construcción de estos modelos no posee una elaboración complicada (Millán 2019).

Mateos del Pino (2009), menciona que el método de Holt es una técnica de alisado exponencial y se utiliza para aquellas series temporales que presentan tendencia y estacionalidad. En este tipo de suavizamiento hace uso de datos históricos para obtener una nueva serie más suave a partir de la cual se hace la predicción. Una de las ventajas del alisado exponencial radica en que se define mediante recurrencias muy simples, de manera que se facilitan los cálculos y se reducen los requerimientos de almacenamiento de datos, esto es de gran ayuda cuando se desarrollan series con grandes datos.

El objetivo que persigue esta etapa es identificar patrones y estimar un modelo estadístico que es la base para generar la información de la muestra, si se desea realizar la predicción de los datos futuros se debe asegurar que las características de la serie son constantes en el tiempo, es decir que la predicción se realizara sobre datos estrictamente estacionarios (Sánchez, 2018).

Validación de los modelos

La longitud de la serie de producción de energía primaria es de $N = 50$, se considerará este valor poblacional en vista de que es el histórico de los datos existentes, la validación de los modelos se efectuará mediante el criterio de Error Medio Absoluto Porcentual (MAPE), el Error Cuadrático Medio (RMSE) y el Criterio de información Bayesiana normalizado (BIC). Las estimaciones se consideran aceptables si se encuentran en el intervalo de confianza del 5%, se ha de seleccionar el modelo que mejor se ajuste con un error mínimo. Una prueba del modelo seleccionado es ver si los residuales estimados a partir de este modelo son de ruido blanco; si lo son, aceptamos el ajuste particular; si no lo son, debemos empezar de nuevo. Por tanto, la metodología es un proceso iterativo (Gujarati & Porter 2010).

Interpretación y comparación de resultados

La construcción de modelos ARIMA para pronosticar es una de las técnicas más empleadas en diferentes ámbitos, gracias a que en corto plazo su capacidad predictiva es

muy cercana a la realidad y la interpretación del modelo emplea entre otras medidas el Error Cuadrático Medio (MSE) y el Error Absoluto Medio (MAE) (Pineda et al., 2017).

De acuerdo con los resultados de Error Medio Absoluto Porcentual (MAPE) y el Error Cuadrático Medio (RMSE) y el Criterio de información Bayesiana normalizado (BIC), se procederá a la interpretación y comparación de resultados, para determinar y seleccionar un modelo que presente las mejores condiciones de diferencia de errores.

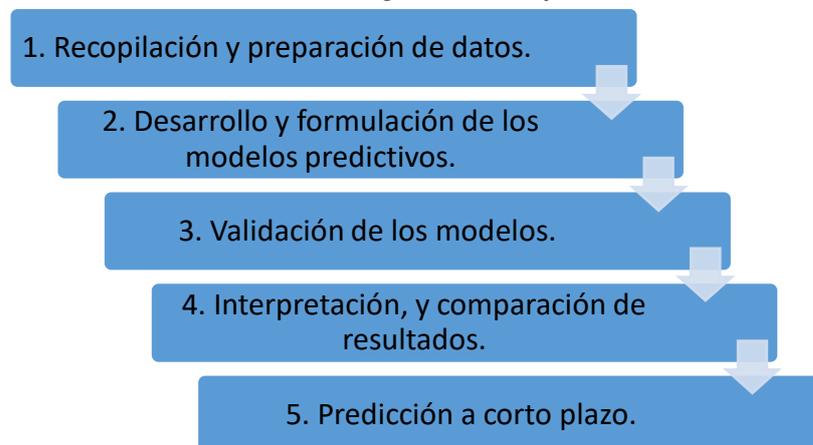
Predicción a corto plazo

Se pretende en este caso utilizar el mejor modelo predictivo seleccionado para realizar una estimación a corto plazo de los siguientes 3 años de la serie estocástica. Un proceso estocástico es un conjunto de variables que se generan de forma aleatoria, además se encuentran ordenadas a la vez equiespaciadas con relación al tiempo, las cuales pueden ser referidas a una o varias características de la variable observable en diferentes momentos (Mauricio, 2007).

Una vez construido el modelo ARIMA su gran aplicación son los pronósticos. En muchos casos, los pronósticos obtenidos por este método son más confiables que los obtenidos de modelos econométricos tradicionales, en particular en el caso de pronósticos de corto plazo (Espino, 2017).

Figura 1

Metodología del trabajo



Resultados

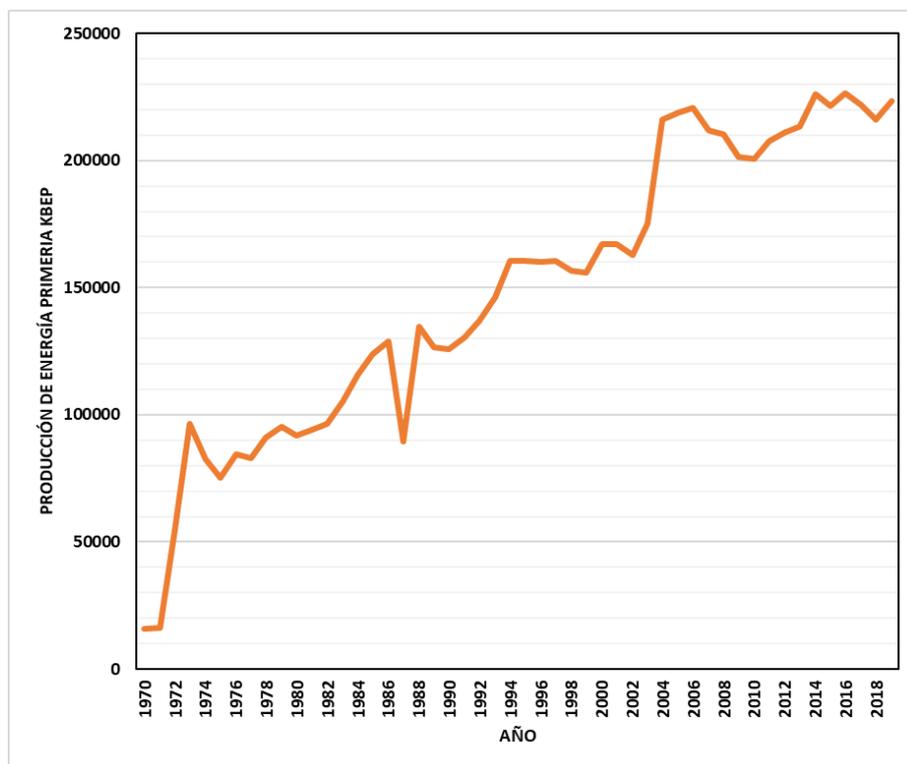
En la primera parte se digitaliza la gráfica de datos históricos con la ayuda del programa webplotdigitizer para obtener datos cuantitativos y así poder obtener la serie temporal de análisis, una vez digitalizado se añaden los datos que se presentan mediante tablas en el Balance Nacional de Energía, para obtener una data de cincuenta observaciones.

Debido a que se han estimado 25 valores de producción de energía primaria a través de la figura que nos ofrece el Balance Nacional de Energía a una tabla de valores numéricos, se procedió a encontrar el error porcentual de la utilización de esta aplicación de libre acceso para la digitalización de imágenes llamada webplotdigitizer, en cuyo caso se encontró un error porcentual promedio de digitalización menor al 1% siendo este igual a 0.48%.

Una vez obtenidos estos valores se procede a tabularlos y graficarlos (figura 2), se observa que se puede describir la presencia de tendencia creciente en la producción de energía primaria, en otras palabras, se puede notar que está presente una pendiente positiva, se percibe también que, en el año 1975, 1988, 2004, se generan saltos que acentúan una línea que termina en un vértice o pico. Las unidades de la producción de energía primaria son kilo barriles equivalentes de petróleo (KBEP).

Figura 2

Serie histórica de producción de energía primaria en Ecuador



Fuente: Balance Energético Nacional (2019)

En relación con la preparación de los datos se inició confeccionando un diagrama de cajas y bigotes, el cual nos permitirá visualizar los candidatos a outliers y los seleccionaremos a través del criterio de Chauvenet (Barrios et al., 2016). De acuerdo con este análisis se

afirma que existen candidatos a outliers o valores atípicos en: tres (3), cuatro (4), dieciocho (18), diecinueve (19), treinta y cinco (35) en total 5 candidatos estos valores son atribuidos a los siguientes años 1972, 1973, 1987, 1988 y 2004. Una vez determinados los candidatos a outliers se procede a la elección mediante el criterio de Chauvenet según este criterio de Puntos Críticos de Chauvenet (P_{ch}) el valor para el número de datos igual a 49 es $P_{ch} = 2.52$, el cual permitirá realizar el contraste con los coeficientes encontrados o calculados y así establecer los outliers o valores atípicos. Si se realiza la comparación para cada valor de la serie temporal se puede observar que existen valores de coeficiente calculado de Chauvenet (P_i) mayores al punto crítico P_{ch} , entonces se afirma que existen outliers o valores atípicos en, cuatro (4), dieciocho (18), diecinueve (19), treinta y cinco (35), descartando el valor de tres (3), estos espacios temporales son atribuidos a los siguientes años 1973, 1987, 1988 y 2004.

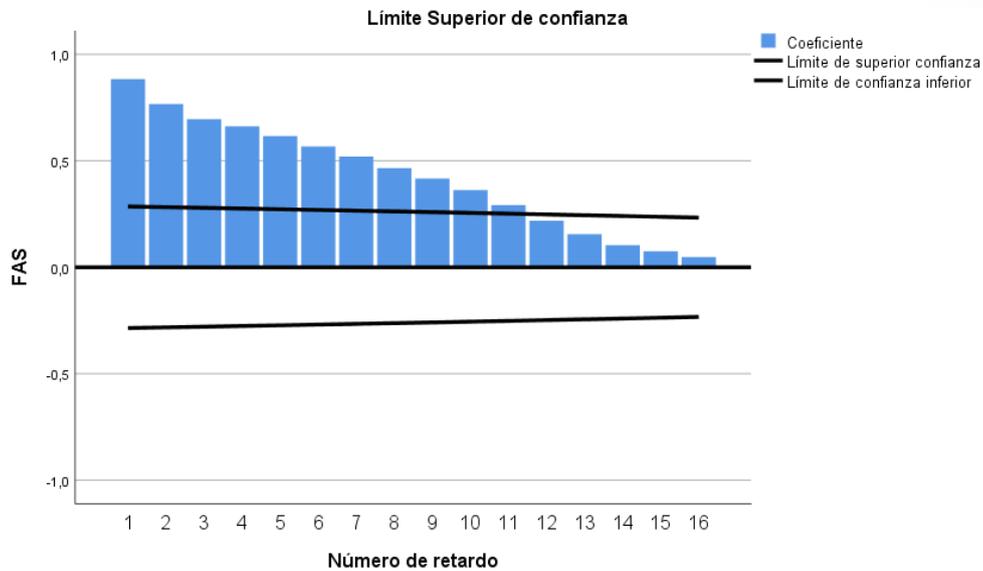
Con la finalidad de que estos valores no degeneren ni distorsionen los estadísticos descriptivos, mediante la ayuda del software estadístico SPSS se los va a modificar con la opción detección automática de valores atípicos y su transformación (Marqués, 2015), los tipos de outliers que permite seleccionar son: aditivo, cambio de nivel, innovador, transitorio, tendencial local y parche aditivo.

Para obtener los diferentes modelos predictivos, se ha trabajado previamente con los datos para no eliminar los outliers y detectarlos automáticamente con el software SPSS y que los transforme o modifique, la segunda acción que se hizo fue transformar la serie de tiempo a una estacionaria, a partir de esta base de datos se determinará la gráfica de autocorrelación simple (FAS) figura 3 y la gráfica de autocorrelación parcial (FAP) figura 4. Vamos a comprobar si estamos ante una serie temporal estacionaria, para ello deberemos ver cómo se comporta a lo largo del año, y si el comportamiento entre periodos del mismo año es el mismo en el resto (García, 2020).

Se evidencia en la gráfica de autocorrelación simple que las barras de coeficientes o barras de significancia muestran una tendencia decreciente suavizada es decir no presentan saltos e interrupciones en cada paso o número de retardo, que comparándolo con las gráficas teóricas de autocorrelación simple determinan que se está en la presencia de un modelo autorregresivo, también se debe hacer la comparación de la gráfica obtenida de autocorrelación parcial con la teórica, la cual nos dará el valor del parámetro del modelo autorregresivo. Si visualizamos con detenimiento la figura 4 se notará que solo existe una barra que sobresale las bandas de confianza lo cual implica que el modelo depende o se explica únicamente con un rezago anterior del tiempo.

Figura 3

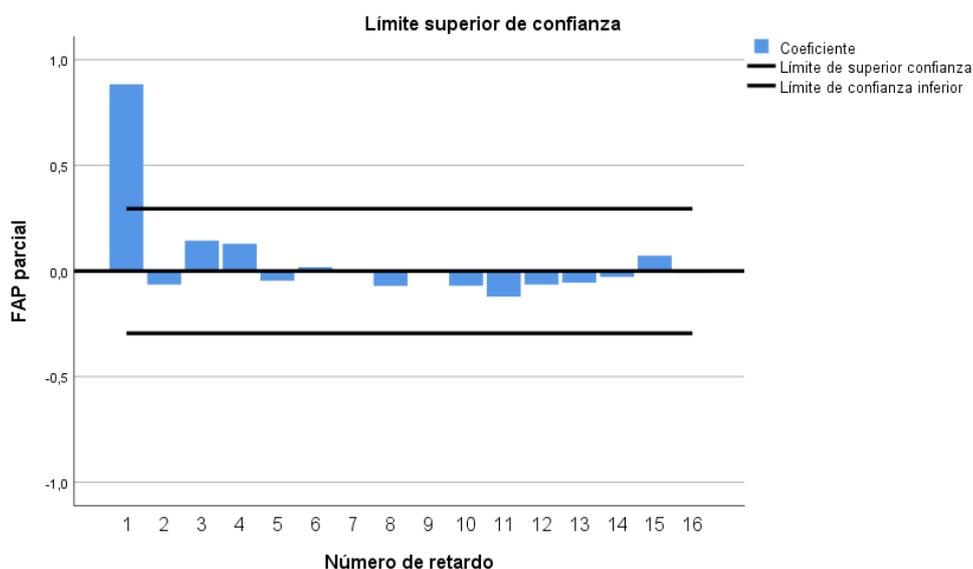
Autocorrelación simple: límite de confianza superior



La autocorrelación parcial indica que el valor de integración o diferencia del modelo ARIMA es $I = 1$, es decir que se conseguirá obtener una serie temporal estacionaria mediante una diferencia de sus valores y de esa manera poder realizar diversas apreciaciones estadísticas.

Figura 4

Autocorrelación parcial: límite de confianza superior



En concordancia con los datos obtenidos de los dos correlogramas se establece en general el modelo lineal auto regresivo como:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_t$$

Donde:

Y_t = Pronóstico para cualquier valor futuro.

α_0 = Constante del modelo

α_1 = Constante del modelo Autorregresivo AR

Y_{t-1} = Valor real anterior para el periodo de tiempo.

ε_t = Error en el instante t.

Como siguiente paso y con la ayuda del software estadístico SPSS se obtienen los valores de los coeficientes: α_0 ; α_1 .

Tabla 1

Parámetros modelo a ARIMA (1,1,0)

| | | Parámetros Modelo_a ARIMA (1,1,0) | | | |
|--------------------------------|------------|-----------------------------------|---------|-------|-------|
| Modelo | | Estimación | SE | t | Sig. |
| Producción de Energía Primaria | Constante | 3365,526 | 962,622 | 3,496 | 0,001 |
| | AR Retardo | 0,074 | 0,147 | 0,507 | 0,615 |
| | 1 | | | | |
| | Diferencia | 1 | | | |

De acuerdo con los parámetros de la tabla 1 ARIMA (1,1,0) obtenemos el siguiente modelo:

$$Y_t = 3365,526 + 0,074 Y_{t-1} + \varepsilon_t$$

Del modelo a descrito se puede mencionar:

- El valor autorregresivo del modelo $AR = 1$, significa que el modelo va a estar en función de datos del pasado anterior Y_{t-1} y del error en el instante ε_t
- La componente de medias móviles $MA = 0$, indica que no se considera el error en el instante pasado ε_{t-1}
- El factor $\alpha_1 \rightarrow 0$ lo que implica que existe una ponderación no muy significativa con el valor anterior para el periodo de tiempo.
- La significancia o el p valor para el constante cumple con: $0,001 < 0,05$ por lo que este término tiene un peso preponderante en el modelo.

Tabla 2

Estadísticos de ajuste del Modelo_a ARIMA (1,1,0)

| Estadísticos de ajuste del Modelo_a ARIMA (1,1,0) | | |
|---|-------|-----------------|
| RMSE | MAPE | BIC normalizado |
| 5998,179 | 3,360 | 18,034 |

De la tabla 2 se evidencia que el error medio absoluto (MAPE) es igual a 3,360 % que es un valor menor que en el intervalo de confianza del 5%.

Tabla 3

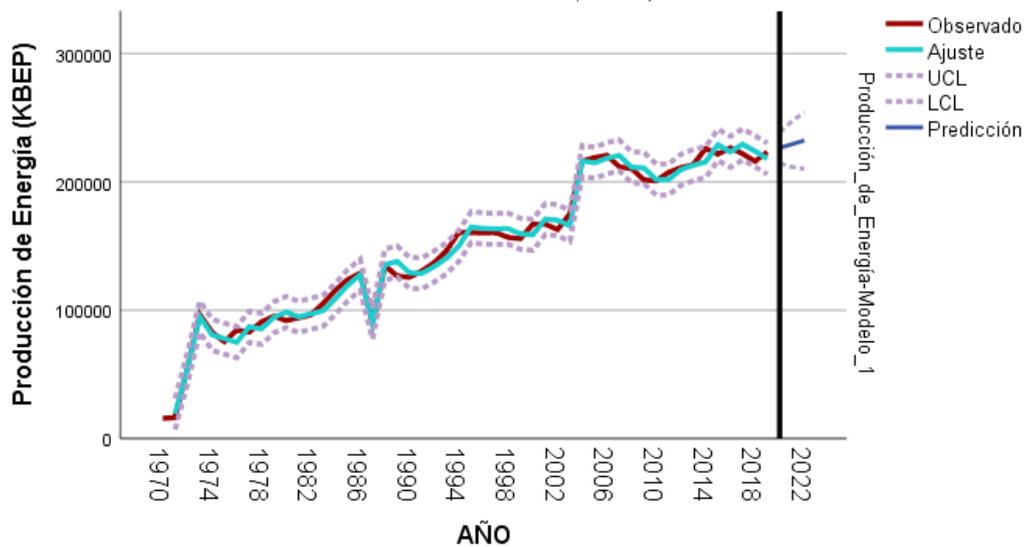
Predicción Modelo_a ARIMA (1,1,0)

| Predicción Modelo_a ARIMA (1,1,0) | | | | |
|-----------------------------------|------------|------------|------------|------------|
| Modelo | | 2020 | 2021 | 2022 |
| Producción de Energía Primaria | Predicción | 226679,447 | 231809,994 | 236940,541 |
| | UCL | 248948,452 | 263303,123 | 275511,589 |
| | LCL | 204410,442 | 200316,865 | 198369,493 |

Si se contempla la tabla 3, se puede decir que la estimación de producción de energía para el año 2022 es de un valor de 236940,541 (KBEP) en un intervalo máximo y mínimo de [275511,589; 198369,493] (KBEP).

Figura 5

Modelo ARIMA (1,1,0)



De la figura 5 se infiere que, la predicción de la producción de energía primaria es creciente con el ajuste mediante Modelo ARIMA (1,1,0), en la primera parte desde el año

1971 hasta 1990 se sobrepone la serie histórica observada y el ajuste, y posteriormente se encuentran cercanas entre las dos, esto indica que el ajuste es aceptable.

A continuación, se procede a variar los parámetros para realizar un modelo ARIMA con diferentes características, se encuentra un modelo ARIMA con (0,1,0).

Tabla 4

Modelo_b ARIMA (0,1,0)

| Modelo | Parámetros Modelo_b ARIMA (0,1,0) | | | | |
|--------------------------------|-----------------------------------|----------|---------|-------|-------|
| | Estimación | SE | t | Sig. | |
| Producción de Energía Primaria | Constante | 3425,813 | 879,185 | 3,897 | 0,000 |
| | Diferencia | 1 | | | |

$$Y_t = 3425,813 + \varepsilon_t$$

Del modelo_b descrito se puede mencionar:

- El valor autorregresivo del modelo $AR = 0$, significa que el modelo no va a depender de los datos del pasado anterior Y_{t-1} .
- La componente de medias móviles $MA = 0$, indica que no se considera el error en el instante pasado ε_{t-1} .
- La significancia o el p valor para el constante cumple con: $0,000 < 0,05$ por lo que este término tiene un peso preponderante en el modelo.

Tabla 5

Estadísticos de ajuste del Modelo_b ARIMA (0,1,0)

| Estadísticos de ajuste del Modelo_b ARIMA (0,1,0) | | |
|---|-------|-----------------|
| RMSE | MAPE | BIC normalizado |
| 5943,322 | 3,419 | 17,936 |

De la tabla 5 se evidencia que el error medio absoluto (MAPE) es igual a 3,419 % que es un valor menor que en el intervalo de confianza del 5%.

Tabla 6

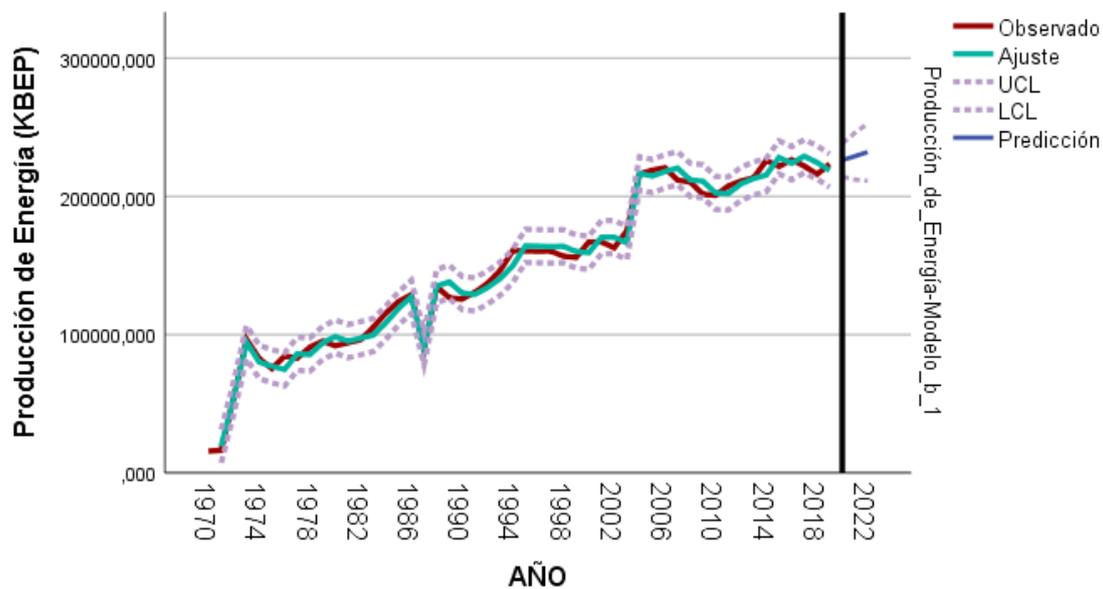
Predicción Modelo b ARIMA (0,1,0)

| Modelo | Predicción Modelo_b ARIMA (0,1,0) | | | |
|--------------------------------|-----------------------------------|------------|------------|------------|
| | | 2020 | 2021 | 2022 |
| Producción de Energía Primaria | Predicción | 226178,361 | 229068,254 | 232027,704 |
| | UCL | 238172,471 | 246030,487 | 252802,111 |
| | LCL | 214184,252 | 212106,022 | 211253,297 |

Si se contempla la tabla 6, se puede decir que la estimación de producción de energía para el año 2022 es de un valor de 232027,704 (KBEP) en un intervalo máximo y mínimo de [252802,111; 211253,297] (KBEP).

Figura 6

Modelo ARIMA (0,1,0)



De la figura 6 se infiere que, la predicción de la producción de energía primaria es creciente con el ajuste mediante Modelo ARIMA (0,1,0), en la primera parte desde el año 1971 hasta 1974 se sobrepone la serie histórica observada y el ajuste, y posteriormente se encuentran cercanas entre las dos, esto indica que el ajuste es aceptable.

El siguiente modelo por obtener es en el cual interviene los promedios móviles (MA), los parámetros a considerar son:

$$Y_t = \mu - \omega_1 \varepsilon_{t-1} + \varepsilon_t$$

Donde:

Y_t = Pronóstico para cualquier valor futuro.

μ = Constante del modelo

ω_1 = Constante del modelo de medias móviles MA

ε_{t-1} = Valor del error anterior para el periodo de tiempo.

ε_t = Error en el instante t.

Tabla 71

Modelo_c ARIMA (0,1,1)

| Modelo | | Parámetros Modelo_c ARIMA (0,1,1) | | | |
|--------------------------------|------------|-----------------------------------|---------|--------|-------|
| | | Estimación | SE | t | Sig. |
| Producción de Energía Primaria | Constante | 3373,909 | 952,627 | 3,542 | 0,001 |
| | Diferencia | 1 | | | |
| | MA | -0,070 | 0,151 | -0,465 | 0,644 |
| | Retardo 1 | | | | |

$$Y_t = 3373,909 + 0,070 \varepsilon_{t-1} + \varepsilon_t$$

Del modelo_c descrito se puede mencionar:

- El valor autorregresivo del modelo $AR = 0$, significa que el modelo no va a depender de los datos del pasado anterior Y_{t-1} .
- La componente de medias móviles $MA = 1$, indica que se debe considerar el error en el instante pasado ε_{t-1} ya que el modelo esta en función de esta variable
- El factor ω_1 es negativa y muy cercana a cero lo que implica que existe una tendencia creciente en el modelo.
- La significancia o el p valor para el constante cumple con: $0,001 < 0,05$ por lo que este término tiene un peso preponderante en el modelo.

Tabla 8

Estadísticos de ajuste del Modelo_c ARIMA (0,1,1)

| Estadísticos de ajuste del Modelo_c ARIMA (0,1,1) | | |
|---|-------|-----------------|
| RMSE | MAPE | BIC normalizado |
| 5999,570 | 3,365 | 18,034 |

De la tabla 8 se evidencia que el error medio absoluto (MAPE) es igual a 3,365 % que es un valor menor que en el intervalo de confianza del 5%.

Tabla 9

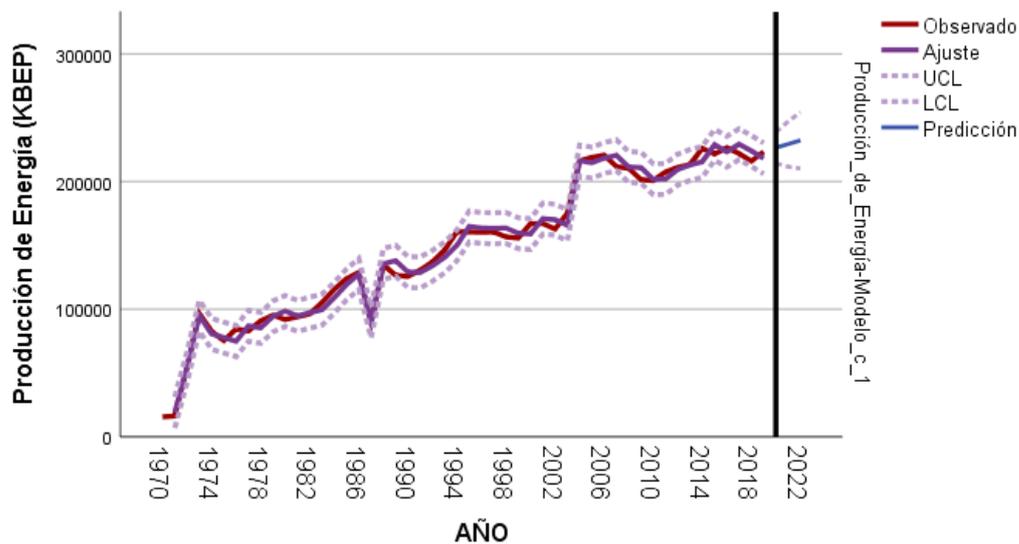
Predicción Modelo_c ARIMA (0,1,1)

| Modelo | | Predicción Modelo_c ARIMA (0,1,1) | | |
|--------------------------------|------------|-----------------------------------|------------|------------|
| | | 2020 | 2021 | 2022 |
| Producción de Energía Primaria | Predicción | 226499,069 | 229342,437 | 232253,897 |
| | UCL | 238615,286 | 247087,285 | 254230,170 |
| | LCL | 214382,853 | 211597,589 | 210277,625 |

Si se contempla la tabla 9, se puede decir que la estimación de producción de energía para el año 2022 es de un valor de 232253,897 (KBEP) en un intervalo máximo y mínimo de [254230,170; 210277,625] (KBEP).

Figura 7

Modelo ARIMA (0,1,1)



De la figura 7 se infiere que, la predicción de la producción de energía primaria es creciente con el ajuste mediante Modelo ARIMA (0,1,1), en la primera parte desde el año 1971 hasta 1974 se sobrepone la serie histórica observada y el ajuste, y posteriormente se encuentran cercanas entre las dos, esto indica que el ajuste es aceptable.

Ahora corresponde encontrar un modelo aplicando suavizamiento exponencial o modelo Holt, la ecuación que la rige es:

$$Y_{t+1} = \alpha Y_t + (1 - \alpha) Y_{t-1}$$

Tabla 10

Modelo_h Suavizamiento Exponencial Holt

| Parámetros Modelo_h Suavizamiento Exponencial Holt | | | | | |
|--|--------------|------------|-------|-------|-------|
| Modelo | | Estimación | SE | t | Sig. |
| Producción de Energía Primaria | Alfa (nivel) | 0,970 | 0,143 | 6,780 | 0,000 |

$$Y_{t+1} = 0,97 Y_t + 0,03 Y_{t-1}$$

Del modelo_h descrito se puede mencionar:

- El valor que influye con mayor ponderación en el modelo va a depender de los datos del pasado anterior Y_t , es decir del inmediato anterior con un valor de 0,97
- La significancia o el p valor para la constante Alpha de suavizamiento con: $0,000 < 0,05$ por lo que este término tiene un peso preponderante en el modelo.
- Los valores atípicos, mediante este método no han sido identificados por el sistema en el software SPSS.

Tabla 11

Estadísticos de ajuste del Modelo_h Holt

| Estadísticos de ajuste del Modelo_h Holt | | |
|--|-------|-----------------|
| RMSE | MAPE | BIC normalizado |
| 14631,702 | 7,769 | 19,260 |

De la tabla 11 se evidencia que el error medio absoluto (MAPE) es igual a 7,769 % que es un valor mayor que en el intervalo de confianza del 5%.

Tabla 12

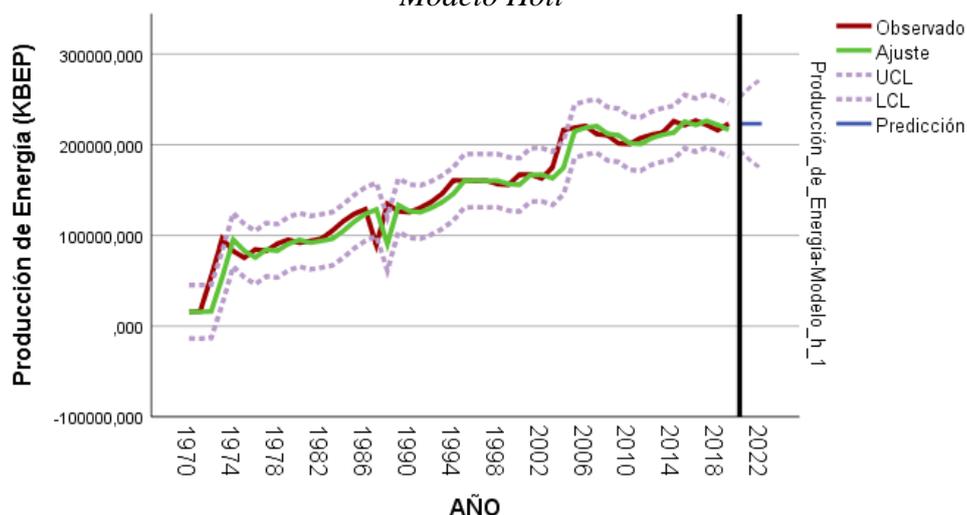
Predicción Modelo_h Holt

| Modelo | | Predicción Modelo_h Holt | | |
|---------------|------------|--------------------------|------------|------------|
| | | 2020 | 2021 | 2022 |
| Producción de | Predicción | 223153,660 | 223153,660 | 223153,660 |
| Energía | UCL | 252557,167 | 264117,847 | 273069,219 |
| Primaria | LCL | 193750,153 | 182189,473 | 173238,102 |

Si se contempla la tabla 12, se puede decir que la estimación de producción de energía para el año 2022 es de un valor de 223153,660 (KBEP) en un intervalo máximo y mínimo de [273069,219; 173238,102] (KBEP).

Figura 8

Modelo Holt



De la figura 8 se infiere que, la predicción de la producción de energía primaria se mantiene constante no es creciente con el ajuste mediante Modelo Holt, en toda la serie no se sobrepone la serie histórica observada y el ajuste siempre se encuentran cercanas entre las dos, esto indica que el ajuste es aceptable.

Tabla 13
Modelos Predictivos

| | RMSE | MAPE | BIC | Expresión |
|------------------------|-----------|-------|--------|--|
| Modelo_a ARIMA (1,1,0) | 5998,179 | 3,360 | 18,034 | $Y_t = 3365,526 + 0,074 Y_{t-1} + \varepsilon_t$ |
| Modelo_b ARIMA (0,1,0) | 5943,322 | 3,419 | 17,936 | $Y_t = 3425,813 + \varepsilon_t$ |
| Modelo_c ARIMA (0,1,1) | 5999,570 | 3,365 | 18,034 | $Y_t = 3373,909 + 0,070 \varepsilon_{t-1} + \varepsilon_t$ |
| Modelo_h Holt | 14631,702 | 7,769 | 19,260 | $Y_{t+1} = 0,97 Y_t + 0,03 Y_{t-1}$ |

A la luz de los resultados se puede afirmar que:

- El error cuadrado medio RMSE se encuentra en un mismo rango en la familia de los modelos ARIMA puesto que los valores atípicos, fueron detectados y remplazados automáticamente por el sistema, mientras que en el modelo Holt no fueron identificados datos atípicos. Además, se debe indicar que el RMSE es menor para el modelo_b
- El error medio absoluto porcentual MAPE, que tiene una bondad de ajuste con menor error es el modelo_a.
- El Criterio de información Bayesiana BIC normalizado es exactamente igual en el modelo_a y modelo_c, pero el menor valor lo obtiene el modelo_b puesto que no depende de componentes autorregresivas y de promedios móviles.

Conclusiones

- Se puede concluir que Modelo_a ARIMA (1,1,0), presenta el menor error medio absoluto porcentual (MAPE) igual 3.36 % que se encuentra dentro de los intervalos de confianza establecidos para los modelos predictivos en un 5%. Si se observan los datos del MAPE de la tabla 13, se verifica que no hay diferencias marcadas entre la familia de los modelos ARIMA, cada uno de los modelos predictivos construidos se encuentran en porcentajes inferiores a 3.419%.
- Se puede afirmar de acuerdo con los datos obtenidos que los modelos predictivos hallados son estrictamente autorregresivos es decir que son métodos iterativos explícitos, puesto que determinan el valor de Y_t en dependencia con el anterior

resultado Y_{t-1} , en el cual no intervienen los residuos de los errores, esto indica que no interviene la componente de medias móviles.

- El modelo que mejor se ajusta a las bondades de predicción de los analizados es Modelo_a ARIMA (1,1,0) cuya expresión es: $Y_t = 3365,526 + 0,074 Y_{t-1} + \varepsilon_t$.
- Se puede estimar en base al modelo predictivo ARIMA (1,1,0), que la energía para el año 2022 en Ecuador podría ser de 236940,541 kilo barriles equivalentes de petróleo (KBEP), con una fluctuación superior e inferior en el intervalo de [275511,589 , 198369,493](KBEP)

Referencias bibliográficas

Alonso, A. M. (2019). Introducción al Análisis de Series Temporales.

Balance Energético Nacional. (2019). Ministerio de recursos renovables y energía. <https://www.recursosyenergia.gob.ec/wp-content/uploads/2020/12/Balance-Energetico-Nacional-2019-1.pdf>

Barrios, R., Castañeda, M., Pedraza, C., Vásquez, J. D. H., & Ibañez, I. (2016). Eliminación de outliers: una estrategia para reducir la incertidumbre tipo a en la calibración de balanzas. In *Simposio de metrología*.

del Pino, Mateo. (2009). Previsión de ventas en una gran cadena de tiendas (Master's thesis, Universitat Politècnica de Catalunya).

Espino Timón, C. (2017). Análisis predictivo: técnicas y modelos utilizados y aplicaciones de este-herramientas Open Source que permiten su uso.

García Corona, J. (2020). Una visión didáctica de modelos predictivos para una serie de demanda eléctrica en España.

González Casimiro, M. P. (2009). Análisis de series temporales: Modelos ARIMA.

Gujarati, D., & Porter, D. (2010). *Econometría*. 673-773, ISBN: 978-607-15-0294-0.

Herrera Granda, D. E. (2019). Predicción de demanda eléctrica mediante la aplicación de modelos ARIMA y SARIMA en lenguaje de programación R—caso de estudio en la Empresa Eléctrica Quito (Bachelor's thesis, Quito, 2019.).

Marqués, M. P. (2015). *Minería de datos: a través de ejemplos*. Alpha Editorial.

Mauricio, J. A. (2007). Análisis de series temporales. Universidad Complutense de Madrid.

Millán Gordo, E. A. (2019). Pronóstico de la demanda de pasajeros aéreos en nueve aeropuertos regionales colombianos.

Peña, D. M. (2010). Análisis de series temporales. Editorial Alianza.

Pineda, S. E. P., Aguilar, J. A. H., & Arroyo-Figueroa, G. (2017). Aplicación de modelos auto regresivos para la predicción de generación de energía eléctrica a partir de datos eólicos. Res. Comput. Sci., 139, 59-70.

Sánchez Sánchez, D. A. (2018). Modelo ARIMA para el pronóstico de la producción de cacao en el Perú 2012-2018.

El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Conciencia Digital**.



El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Conciencia Digital**.



Indexaciones

