

MODELO A CORTO PLAZO PARA EL PRONÓSTICO DEL PRECIO DE BOLSA
DEL MERCADO DE ENERGÍA ELÉCTRICA EN COLOMBIA

MICHAEL ANGELO SANTANA MONTES
BRAYAN FERNANDO CUELLAR MACIAS

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍAS
ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
BUCARAMANGA

2022

MODELO A CORTO PLAZO PARA EL PRONÓSTICO DEL PRECIO DE BOLSA
DEL MERCADO DE ENERGÍA ELÉCTRICA EN COLOMBIA

MICHAEL ANGELO SANTANA MONTES
BRAYAN FERNANDO CUELLAR MACIAS
*Trabajo de Grado para optar al título de
Ingeniero Electricista*

Director

IVÁN DAVID SERNA SUÁREZ

Ph.D en Ingeniería

Codirectores

CARLOS ALFONSO MANTILLA DUARTE

M.Sc en estadística aplicada

OSCAR ARNULFO QUIROGA QUIROGA

Ph.D en Tecnología

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍAS
ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
BUCARAMANGA

2022

CONTENIDO

	pág.
Introducción	9
1. Objetivos	13
2. Modelos de pronóstico	14
2.1. Modelo Autorregresivo Integrado de Media Móvil	14
2.1.1. Modelos autorregresivos de orden p - AR(p)	14
2.1.2. Proceso de integración de orden d - I(d)	15
2.1.3. Modelos de medias móviles de orden q - MA(q)	15
2.1.4. Modelos ARIMA(p,d,q)	16
2.2. Modelo de regresión lineal	17
2.3. Modelo de regresión lineal dinámico	18
3. Análisis y modelado de las series de tiempo	19
3.1. Datos y metodología	19
3.2. Implementación del modelo ARIMA	25
3.3. Implementación del modelo de regresión dinámica	27
4. Elaboración de pronósticos	29
5. Conclusión	38
Bibliografía	39

LISTA DE FIGURAS

	pág.
Figura 1. Capacidad efectiva neta del SIN de Colombia. Fuente: Adaptado de XM. PARATEC - Parámetros técnicos del SIN.	10
Figura 2. Ejemplo de modelo de regresión lineal simple. Fuente: Adaptado de Rob J Hyndman y George Athanasopoulos. Forecasting: principles and practice.	17
Figura 3. Diagrama de flujo de la metodología utilizada. Fuente: Autores.	20
Figura 4. Serie de tiempo de los aportes de energía hidráulica en Colombia. Fuente: Autores.	21
Figura 5. Descomposición de la serie de tiempo de los aportes de energía hidráulica en Colombia por año. Fuente: Autores.	22
Figura 6. Serie de tiempo del precio de bolsa promedio del mercado de energía eléctrica en Colombia. Fuente: Autores.	23
Figura 7. Descomposición de la serie de tiempo del precio de bolsa promedio del mercado de energía en Colombia por año. Fuente: Autores.	24
Figura 8. Serie de tiempo del precio bolsa promedio - Ventana seleccionada para la evaluación de los modelos. Fuente: Autores.	29
Figura 9. Pronóstico del precio de bolsa promedio con ventana de predicción de un día. Fuente: Autores.	31
Figura 10. Pronóstico del precio de bolsa promedio con ventana de predicción de cuatro días. Fuente: Autores.	32
Figura 11. Pronóstico del precio de bolsa promedio con ventana de predicción de siete días. Fuente: Autores.	33

Figura 12. Valores de error medio absoluto de cada pronóstico. Fuente: Autores.	35
Figura 13. Valores de error porcentual medio absoluto de cada pronóstico. Fuente: Autores.	36
Figura 14. Pronóstico del precio de bolsa promedio con ventana de predicción de un día eliminando la componente de desfase. Fuente: Autores.	37

LISTA DE TABLAS

	pág.
Tabla 1. Criterios AICc de los modelos de regresión tentativos. Fuente: Autores.	28
Tabla 2. Error máximo y mínimo cometido en el pronóstico con la técnica ARIMA. Fuente: Autores.	34
Tabla 3. Error máximo y mínimo cometido en el pronóstico con la técnica de regresión dinámica. Fuente: Autores.	34

RESUMEN

TÍTULO: Modelo a corto plazo para el pronóstico del precio de bolsa del mercado de energía eléctrica en Colombia *

AUTORES: Michael Angelo Santana Montes - Brayán Fernando Cuellar Macías **

PALABRAS CLAVE: Aportes hídricos, arima, mercado de energía, precio de bolsa, pronóstico, regresión dinámica, series de tiempo.

DESCRIPCIÓN:

En el mercado de energía eléctrica de Colombia, la predicción confiable del precio de bolsa en el corto plazo representa un papel fundamental en las estrategias de planeación y licitación de los agentes participantes para poder maximizar sus beneficios en la dinámica del mercado. En este documento se desarrollaron dos metodologías a fin de pronosticar el valor promedio del precio de bolsa a partir del enfoque de las series de tiempo, utilizando como técnicas de modelado el método ARIMA y un método de regresión dinámica que involucra la serie de los aportes hídricos en calidad de variable predictora. El conjunto de datos empleado contiene la información de las series de tiempo de los aportes hídricos y del precio de bolsa en el periodo comprendido entre el año 2016 y 2020, la elaboración de las predicciones se desarrolló a partir de una de ventana entrenamiento móvil con duración de un año. Los resultados muestran que la técnica ARIMA es ligeramente superior a la de regresión dinámica, sin embargo, la magnitud de los errores de predicción de los modelos se encuentra muy cercanas, demostrando que la inclusión de los aportes hídricos como variable predictora puede mejorar la precisión de los pronósticos.

* Trabajo de grado

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Iván David Serna Suárez, Codirector: Carlos Alfonso Mantilla Duarte. Codirector: Oscar Arnulfo Quiroga Quiroga

ABSTRACT

TITLE: Short-term model for the forecast of the stock market price of the electric power market in Colombia *

AUTHORS: Michael Angelo Santana Montes - Brayan Fernando Cuellar Macias **

KEYWORDS: Hydraulic inputs, arima, energy market, stock market price, forecast, dynamic regression, time series.

DESCRIPTION:

In the Colombian electric power market, the reliable prediction of the exchange price in the short term represents a fundamental role in the planning and bidding strategies of the participating agents in order to maximize their profits in the market dynamics. In this paper two methodologies were developed to forecast the average value of the stock market price from the time series approach, using as modeling techniques the ARIMA method and a dynamic regression method that involves the series of hydraulic inputs as a predictor variable. The dataset used contains the information of the time series of the hydraulic inputs and the stock market price in the period from 2016 to 2020, the elaboration of the predictions was developed from a one-year moving training window. The results show that the ARIMA technique is slightly superior to the dynamic regression technique, however, the magnitude of the prediction errors of the models is very close, demonstrating that the inclusion of hydraulic inputs as a predictor variable can improve the accuracy of the forecasts.

* Bachelor Thesis

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Iván David Serna Suárez, Codirector: Carlos Alfonso Mantilla Duarte. Codirector: Oscar Arnulfo Quiroga Quiroga

Introducción

A partir de la reestructuración realizada en la década de los noventa a fin de mejorar la calidad, disponibilidad y confiabilidad del sector energético, el mercado de energía eléctrica colombiano se ha caracterizado por promover la libre participación de inversionistas privados en los sectores de generación y comercialización¹. La introducción de la competencia en el sector de la generación contribuye a que los precios de bolsa reflejen el costo de la producción de energía². En el caso de Colombia, uno de los factores que más afecta el precio de bolsa es la disponibilidad de los recursos hídricos, debido a que actualmente en el país el 67.3 % de la energía eléctrica generada es producida a partir de las centrales hidroeléctricas³.

Tal como se muestra en la **Figura 1**, la generación restante se divide en centrales térmicas (30.8 %) y centrales de energías renovables (1.94 %). Esta última está conformada por plantas cogeneratoras, energía solar y energía eólica, con porcentajes de 0.812 %, 0.656 % y 0.105 % respectivamente⁴.

¹ BELLO RODRÍGUEZ, Sandra Patricia y BELTRÁN AHUMADA, Robert Baudilio. Caracterización y pronóstico del precio spot de la energía eléctrica en Colombia. En: Revista de la Maestría en Derecho Económico 6.6 (2013), págs. 293-316.

² BOTERO BOTERO, Sergio y CANO CANO, Jovan Alfonso. Análisis de series de tiempo para la predicción de los precios de la energía en la bolsa de Colombia. En: Cuadernos de Economía 27.48 (2008).

³ XM. PARATEC - Parámetros técnicos del SIN. [En línea] Disponible en: <http://paratec.xm.com.co/paratec/SitePages/Default.aspx>. 2020.

⁴ Ibid.

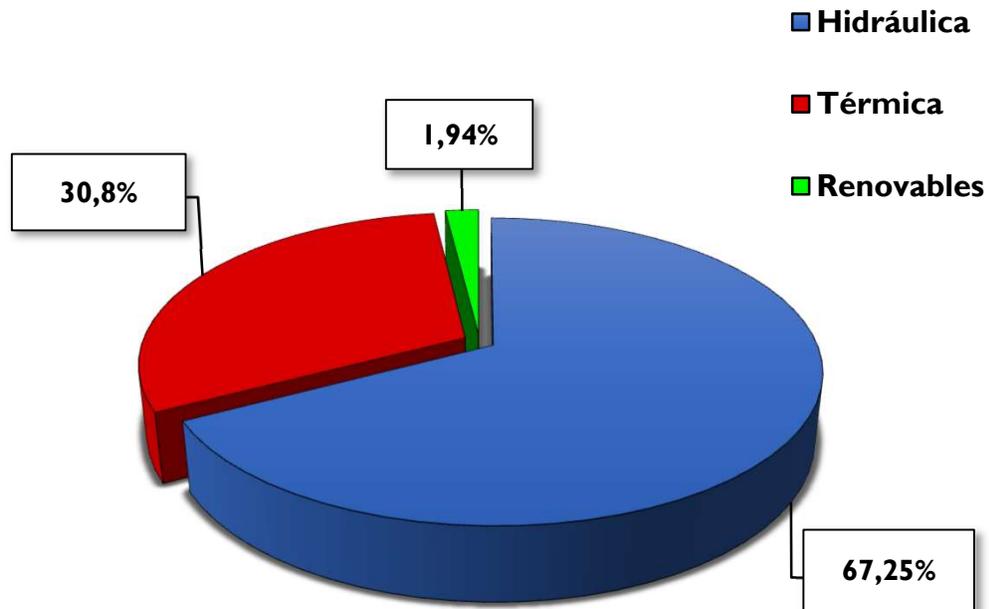


Figura 1. Capacidad efectiva neta del SIN de Colombia.

Fuente: Adaptado de XM. PARATEC - Parámetros técnicos del SIN.

En consecuencia, la desregularización del mercado, la alta dependencia de la generación hidráulica y la ausencia de tecnologías para almacenar energía eléctrica a gran escala pueden llegar a producir incrementos abruptos en el precio de bolsa, convirtiendo su estimación en un gran reto para todos los agentes interesados⁵.

En los últimos años, en diferentes mercados del mundo se han desarrollado gran variedad de métodos y modelos para la predicción del precio de la energía eléctrica⁶, demostrando que aquellos que se encuentran basados en el análisis de la información de las series de tiempo presentan una buena estimación de los valores a futuro en el corto plazo. En el enfoque de las series de tiempo, los modelos autorregre-

⁵ MCHUGH, Catherine, et al. Forecasting Day-ahead Electricity Prices with A SARIMAX Model. En: 2019 IEEE Symposium Series on Computational Intelligence (SSCI). 2019, págs. 1523-1529.

⁶ GAO, G; LO, Fulin y FAN, Fulin. Comparison of ARIMA and ANN models used in electricity price forecasting for power market. En: Energy and Power Engineering 09 (2017), págs. 120-126.

sivos integrados de medias móviles (**ARIMA**) han obtenido resultados significativos y confiables en el área de la predicción de los precios diarios, debido a que permiten manejar los errores causados por autocorrelación y relacionar los valores del presente con los del pasado⁷. En CONTRERAS⁸, los autores aplican la metodología ARIMA basándose en el análisis de las series de tiempo para pronosticar los precios horarios del día siguiente en los mercados eléctricos de España y California, obteniendo errores medios de predicción del 10 % y del 5 % para cada mercado. Por otra parte, en GAO⁹, utilizando la información del mercado del Reino Unido, efectúan la comparación de los modelos ARIMA con una red neuronal artificial (ANN) evaluando su rendimiento en la predicción a corto plazo, demostrando a partir de los errores de pronóstico que el modelo ARIMA especificado supera al modelo de la red neuronal estudiado. Así mismo, en BISSING¹⁰ los autores proponen un modelo híbrido combinando los modelos ARIMA con un modelo de regresión lineal múltiple para mejorar los pronósticos, concluyendo que a partir de la combinación de estos dos modelos es posible obtener un pronóstico confiable de magnitud y forma adecuada. Por esta razón, considerando la importancia que representa el precio spot de la energía eléctrica en la planeación de los ingresos de los agentes participantes en el mercado, en las estrategias de licitación de los productores y en la negociación de contratos a largo plazo¹¹, se propone en el presente trabajo dos metodologías

⁷ MCHUGH. Op. Cit.

⁸ CONTRERAS, J, et al. ARIMA models to predict next-day electricity prices. En: IEEE Transactions on Power Systems 18.3 (2003), págs. 1014-1020.

⁹ GAO. Op. Cit.

¹⁰ BISSING, Daniel, et al. A Hybrid Regression Model for Day-Ahead Energy Price Forecasting. En: IEEE Access 7 (2019), págs. 36833-36842.

¹¹ CONTRERAS. Op. Cit.

para la predicción del precio de bolsa a través del análisis de las series de tiempo, a fin de reducir el riesgo de inversión para garantizar las utilidades de los agentes de generación y consumidores a gran escala del mercado.

En particular, para generar pronósticos a corto plazo del precio de bolsa del mercado de energía eléctrica colombiano, en este trabajo se pretende aprovechar las relación existente entre el precio de bolsa y los aportes hídricos. Para esto es fundamental hacer un reconocimiento de los factores que se deben tener en cuenta en el momento de elegir un modelo para el pronóstico, puesto que su rendimiento estará influenciado no solo por los métodos, sino también, por cómo estos se aplican al problema estudiado. Además, debido a que no existe un método universal que funcione de manera óptima para todos los escenarios, este se deberá elegir conforme a las características y necesidades del mercado¹².

El resto del documento se encuentra organizado de la siguiente manera, en el **capítulo II** se describe el fundamento teórico sobre el cual trabajan las metodologías de pronóstico a comparar, en el **capítulo III** se detalla el procedimiento para la estimación de los modelos, en el **capítulo IV** se presentan los resultados y en el **capítulo V** se suministran las conclusiones.

¹² JIANG, LianLian y HU, Guoqiang. A Review on Short-Term Electricity Price Forecasting Techniques for Energy Markets”. En: 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV). 2018, págs. 937-944.

1. Objetivos

Objetivo general

- Caracterizar el precio de bolsa de la energía eléctrica a corto plazo, a partir de la correlación entre los recursos hídricos y el histórico del precio de bolsa del mercado.

Objetivos específicos

- Analizar las series de tiempo de los recursos hídricos y del precio de bolsa del mercado de energía.
- Caracterizar la correlación entre las series de tiempo de la hidrología y el precio de bolsa del mercado.
- Aplicar el modelo autorregresivo de media móvil (ARIMA) a las series de tiempo, para pronosticar a corto plazo el precio de bolsa del mercado.

2. Modelos de pronóstico

2.1. Modelo Autorregresivo Integrado de Media Móvil

El enfoque de estos modelos auto predictivos con énfasis en las propiedades probabilísticas o estocásticas de las series de tiempo, se fundamenta en predecir los valores de una variable en función de su información pasada, sin la necesidad de incluir otros datos adicionales de variables auxiliares o relacionadas¹³. Es decir, en estos modelos la variable dependiente o pronosticada se encuentra expresada por los datos pasados de sí misma y por los coeficientes o términos estocásticos del error¹⁴. Este tipo de modelos son muy utilizados en el área de la predicción debido a que poseen la capacidad de desarrollar pronósticos con un excelente grado de precisión (tanto para series de tiempo estacionarias y no estacionarias) basándose en una descripción de los patrones históricos de los datos¹⁵.

2.1.1. Modelos autorregresivos de orden p - AR(p) Este tipo de modelos desarrollan el pronóstico de la variable de interés a partir de una combinación lineal de los valores pasados de la serie¹⁶. El orden del modelo (denotado por la letra **p**) hace referencia al número de retardos que se deben incluir para desarrollar las prediccio-

¹³ JIMÉNEZ NAVIA, Breyner y MORENO BUITRAGO, Carlos Adolfo. Modelos estadísticos tipo ARIMA para el pronóstico de incautaciones de drogas ilegales por parte de la armada nacional de Colombia. Bogotá, 2017. Fundación universitaria los libertadores.

¹⁴ MUÑOZ SANTIAGO, Alberto, et al. Pronóstico del precio de la energía en Colombia utilizando ARIMA con IGARCH. En: Revista de Economía del Rosario, 20, (2017); págs. 127-161.

¹⁵ HANKE, John E y WICHERN, Dean W. Pronósticos en los negocios. 9 ed. por Antonio P Enriquez Brito, Ofelia Vizcaíno-Díaz y José Antonio Rosas-Ferrer. Pearson Educación, 2010, págs. 1-576.

¹⁶ HYNDMAN J, Rob y ATHANASOPOULOS, George. Forecasting: principles and practice. 2 ed. [En línea] Disponible en: <https://otexts.com/fpp2/>. OTexts: Melbourne, Australia, 2018.

nes¹⁷. Por lo tanto, el modelo matemático de un proceso autorregresivo de orden p se describe de la siguiente manera:

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

Donde ϕ_p son los coeficientes autorregresivos del modelo, C es una constante y ε_t es un ruido blanco.

2.1.2. Proceso de integración de orden d - I(d) El término d del proceso de integración, hace referencia al número de diferencias que se deben aplicar a una serie de tiempo para que esta sea estacionaria. En consecuencia, una serie de tiempo es integrada de orden d , cuando sus d diferencias son estacionarias¹⁸.

Por ejemplo, la expresión matemática de una serie de tiempo diferenciada dos veces, y por tanto, con un proceso de integración de orden 2 es la siguiente:

$$y'_t = y_t - y_{t-1} \quad (2)$$

$$y''_t = y'_t - y'_{t-1} \quad (3)$$

Donde y'_t y y''_t representan la primera y segunda diferencia de la serie.

2.1.3. Modelos de medias móviles de orden q - MA(q) Los procesos de medias móviles desarrollan la predicción de la variable de interés a partir de los errores de pronósticos pasados del modelo. Es decir, en este tipo de procesos la variable de interés se puede expresar como un promedio móvil ponderado de los últimos errores

¹⁷ FERNÁNDEZ, Victoria. ARIMA. [En línea] Disponible en: <https://rpubs.com/vicferu/394457>. 2018.

¹⁸ MUÑOZ SANTIAGO. Op. Cit.

del pronóstico¹⁹. El orden del modelo (denotado por la letra **q**) determina el número de términos de error que se deben incluir para el pronóstico²⁰.

La expresión matemática de un proceso de medias móviles de orden **q** se presenta a continuación:

$$y_t = C + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q} \quad (4)$$

Donde θ_q son los coeficientes de medias móviles del modelo, C es una constante y ε_t es un ruido blanco.

2.1.4. Modelos ARIMA(p,d,q) La combinación de los procesos de autorregresión, integración y de medias móviles, dan origen a un modelo ARIMA²¹ de orden (p,d,q) que posee la siguiente expresión matemática:

$$y_t^d = C + \phi_1y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q} + \varepsilon_t \quad (5)$$

Donde y_t^d es la serie de tiempo diferenciada. En la medida que se necesite trabajar con modelos de orden superior, resulta mas conveniente utilizar la ecuación del modelo con la notación del operador de retroceso (**B**)²².

$$\underbrace{(1 - \phi_1 B - \dots - \phi_p B^p)}_{AR(p)} \underbrace{(1 - B)^d}_{(d) \text{diferencias}} y_t = C + \underbrace{(1 + \theta_1 B + \dots + \theta_q B^q)}_{MA(q)} \varepsilon_t \quad (6)$$

¹⁹ HYNDMAN. Op. Cit.

²⁰ FERNÁNDEZ. Op. Cit.

²¹ El termino ARIMA se refiere a un acrónimo de la expresión en inglés **Auto Regresive Integrated Moving Average** y hace referencia a cada una de las partes que conforman el modelo, *AR* de autorregresivo, *I* de integrado y *MA* de medias móviles.

²² HYNDMAN. Op. Cit.

2.2. Modelo de regresión lineal

El término de regresión hace referencia a un proceso estadístico que permite matemáticamente cuantificar una relación funcional entre dos o más variables²³. En el caso de la regresión lineal, se fundamenta en estimar la dinámica de la serie de tiempo de la variable interés a partir de una relación lineal con respecto a una variable predictora²⁴.

La expresión matemática de un modelo de regresión lineal con un solo predictor es la siguiente:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (7)$$

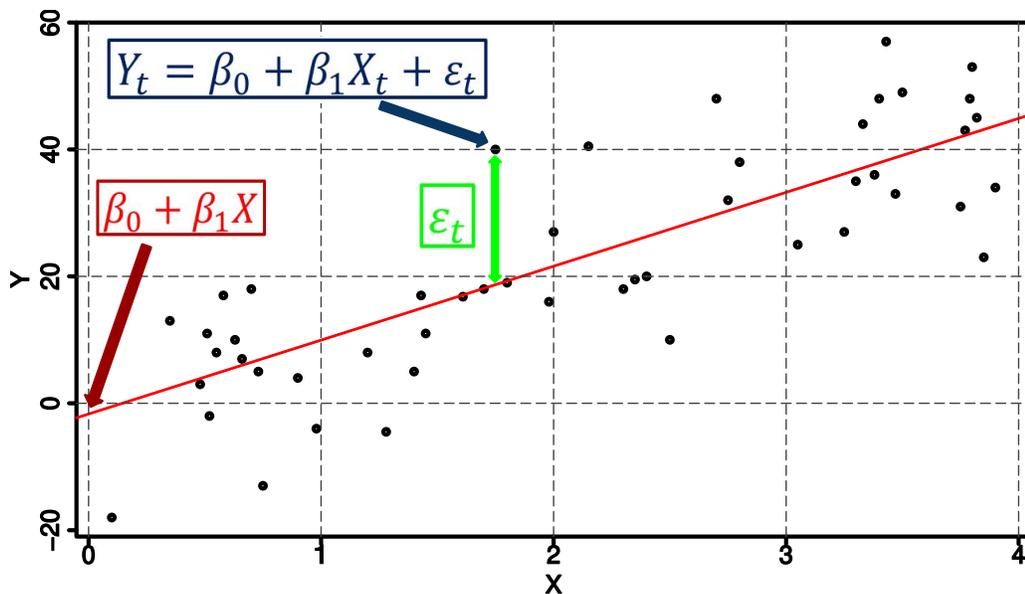


Figura 2. Ejemplo de modelo de regresión lineal simple.

Fuente: Adaptado de Rob J Hyndman y George Athanasopoulos. *Forecasting: principles and practice*.

²³ MONTERO GRANADOS, Roberto. Modelos de regresión lineal múltiple. Universidad de Granada, 2016, págs. 1-61.

²⁴ HYNDMAN. Op. Cit.

En la **Figura 2** se muestra un ejemplo artificial de unos datos modelados a través de este método. Como se observa en la gráfica, el método realiza la mejor estimación de una ecuación lineal para capturar la evolución de la dinámica de los datos. En este, β_0 representa la intercepción con la variable predictora cuando esta toma un valor de cero, β_1 la pendiente de la función y ε_t un error aleatorio incluido por la dispersión de los datos²⁵.

2.3. Modelo de regresión lineal dinámico

La metodología ARIMA permite modelar el comportamiento de una serie de tiempo a partir de la de información pasada de la serie, pero su aplicación no admite la introducción de variables relacionadas que puedan ayudar a mejorar las estimaciones. Por otra parte, los modelos de regresión lineal aceptan la inclusión de muchas variables predictoras, pero carecen de herramientas para manejar la información histórica de los datos. Por esto, a fin de permitir la introducción de información de otros modelos en modelos ARIMA, se puede plantear un modelo de regresión cuyos errores esten modelados por medio de ARIMA²⁶.

Las ecuaciones del modelo serían:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \eta_t, \quad (8)$$

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d \eta_t = (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \quad (9)$$

Donde η_t es el error de la regresión lineal, el cual se modela por medio de ARIMA y ε_t es un ruido blanco que corresponde al error del modelo ARIMA.

²⁵ Ibid.

²⁶ Ibid.

3. Análisis y modelado de las series de tiempo

3.1. Datos y metodología

En este artículo de investigación se proponen dos alternativas con el propósito de pronósticar a corto plazo el precio de bolsa promedio del mercado de energía eléctrica de Colombia. La primera de ellas utiliza la metodología **ARIMA** para generar un modelo de predicción a partir de la serie de tiempo del precio de bolsa, mientras que la segunda emplea una técnica de **regresión dinámica** con la intención de modelar la relación existente entre las series de tiempo del precio de bolsa y los aportes hídricos. En la figura **Figura 3** se describe el diagrama de flujo de las metodologías desarrolladas.

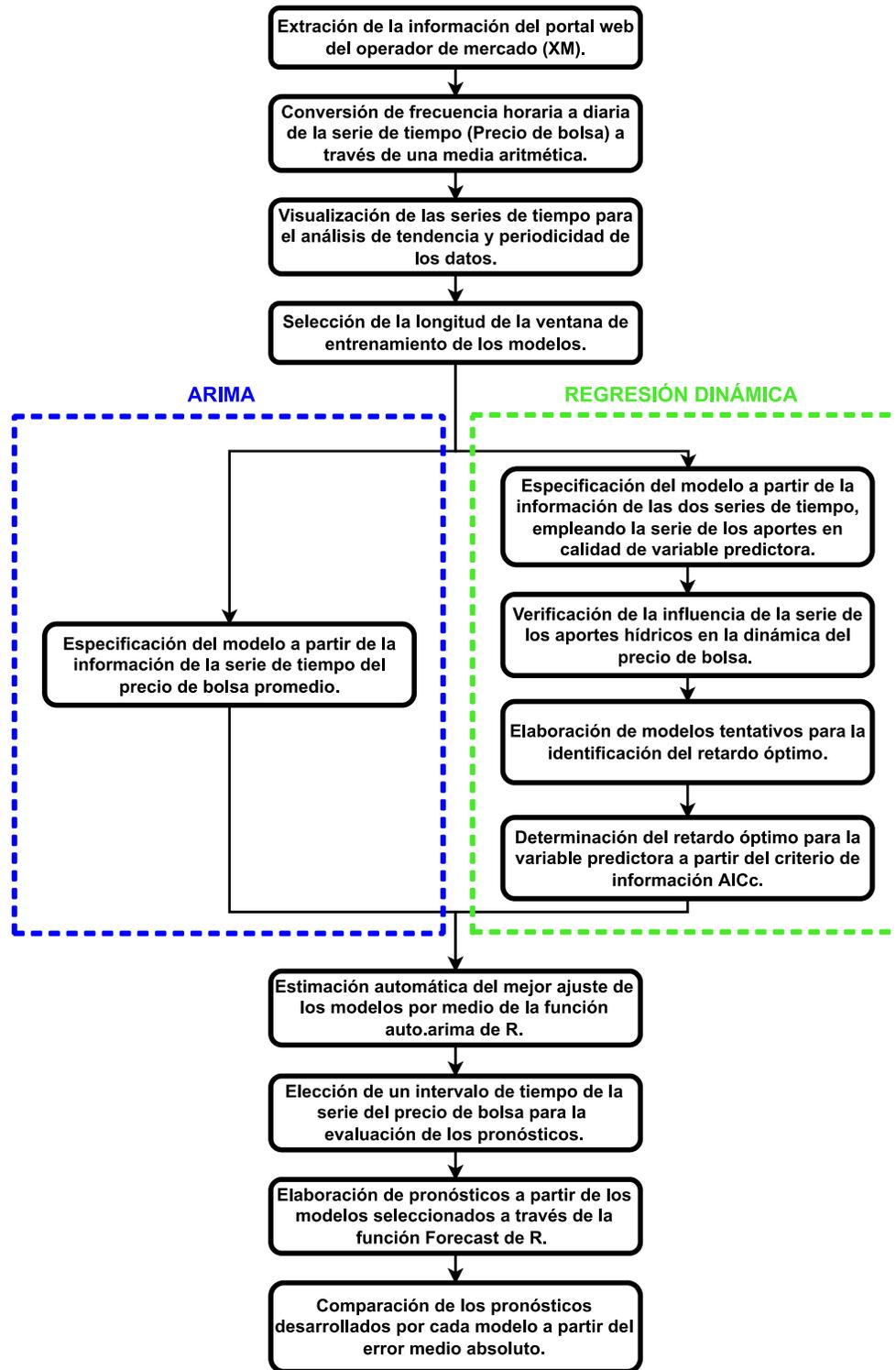


Figura 3. Diagrama de flujo de la metodología utilizada.

Fuente: Autores.

En ambos escenarios se utilizó la función **auto.arima** proporcionada por el paquete **forecast** del **lenguaje de código abierto R** para identificar de manera automática el modelo que mejor se ajusta a cada ventana de tiempo²⁷.

La información utilizada en el procesamiento se obtuvo a través del portal dispuesto por el operador del mercado para la consulta pública de los datos (<https://sinergox.xm.com.co/Paginas/Home.aspx>). Este contiene los valores obtenidos en la dinámica del mercado para las variables estudiadas durante el periodo comprendido entre los años 2016 y 2020. Debido a que las series de tiempo originales (aportes hídricos y precio de bolsa) poseen frecuencias distintas (diaria y horaria), se realizó una media aritmética con los datos del precio de bolsa a fin de obtener una nueva serie con 1825 observaciones de frecuencia diaria.

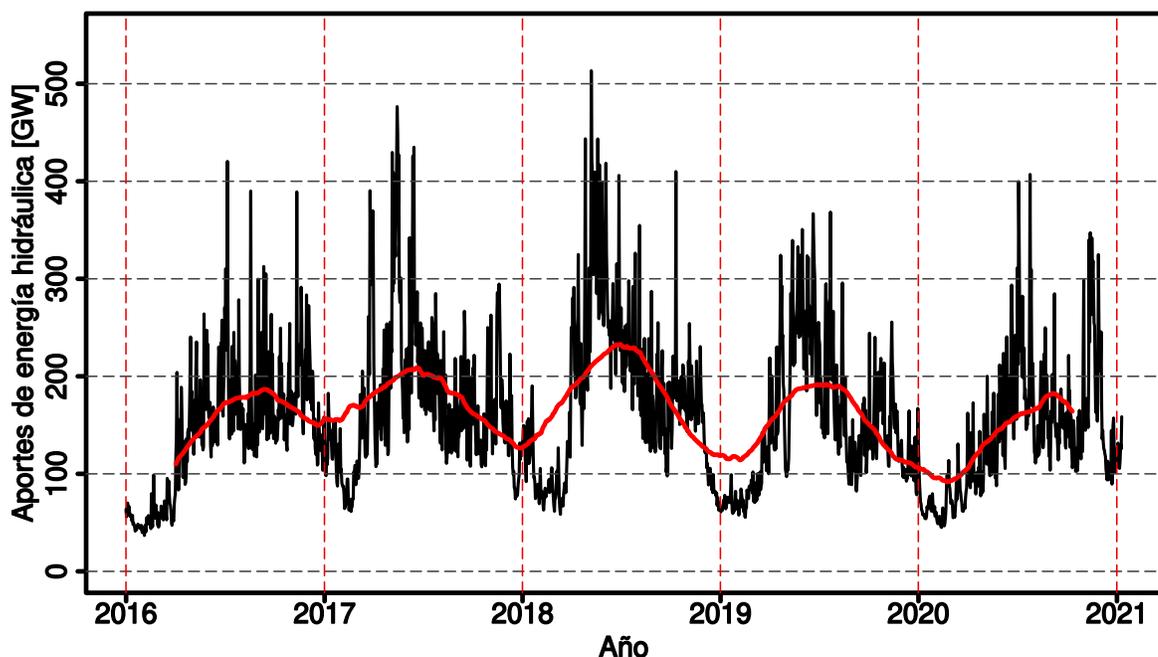


Figura 4. Serie de tiempo de los aportes de energía hidráulica en Colombia.

Fuente: Autores.

²⁷ HUERTAS, Nicolás y DONADO, Santiago. Evaluación crítica de la función `auto.arima` de R en la identificación de procesos de series de tiempo. 2018, págs 1-33. Universidad de los Andes.

En la **Figura 4** se aprecia la serie de tiempo con observaciones diarias obtenida para los aportes hídricos en el periodo comprendido entre el 01 de enero del 2016 y el 31 de diciembre del 2020. Al desarrollar un análisis de su comportamiento es posible identificar a través de la curva roja que en términos de tendencia la serie de tiempo describe una dinámica similar en cada uno de los años.

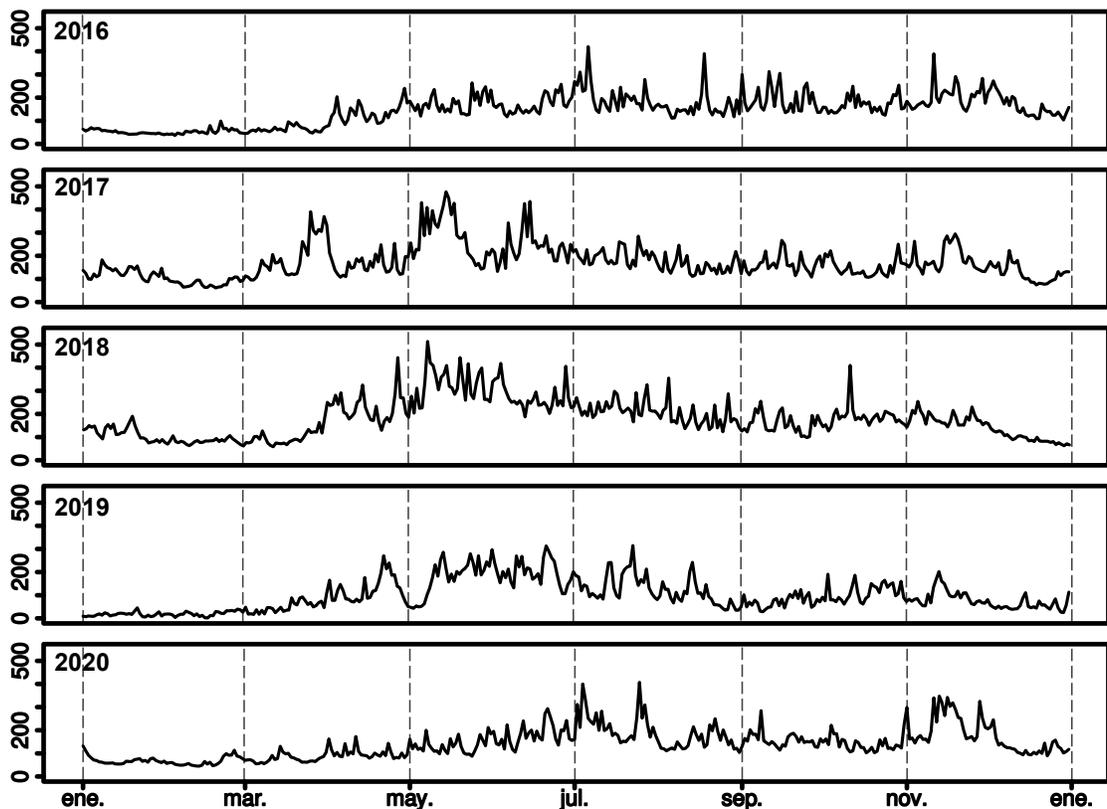


Figura 5. Descomposición de la serie de tiempo de los aportes de energía hidráulica en Colombia por año.

Fuente: Autores.

También, en la **Figura 5** se puede notar, que en los primeros meses se contemplan los valores más bajos de la serie y a partir del tercer mes se presenta un incremento que se sostiene hasta mediados del año, en donde inicia a disminuir manifestando nuevamente un aumento en los dos últimos meses.

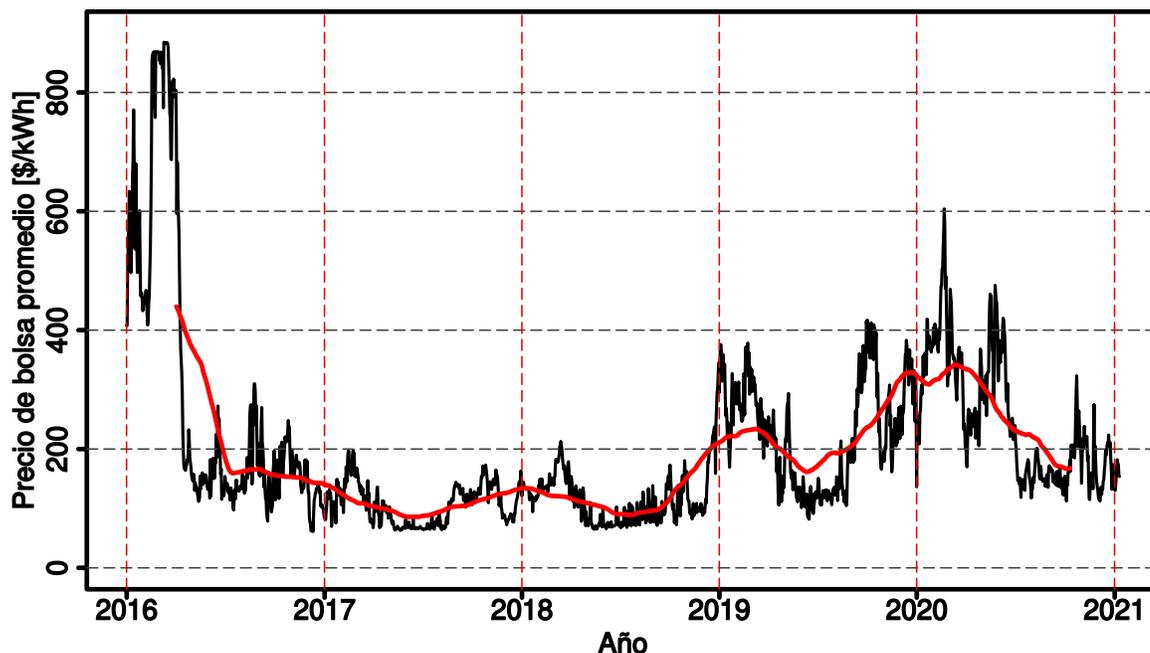


Figura 6. Serie de tiempo del precio de bolsa promedio del mercado de energía eléctrica en Colombia.

Fuente: Autores.

Por otra parte, en la **Figura 6**, se visualiza la serie de tiempo de frecuencia diaria obtenida para el precio de bolsa al calcular la media aritmética con los datos horarios; al igual que la serie de tiempo de los aportes hídricos, los datos se encuentran observados en el mismo intervalo de tiempo. Al estudiar la conducta de la serie se observa que esta presenta un comportamiento especial en los primeros meses del año 2016, debido a que en esa época el país se encontraba atravesando un fuerte episodio del fenómeno del niño que ocasionó ausencia de lluvias y sequías a lo largo del territorio²⁸. Sin embargo, como se observa en la **Figura 7**, a pesar de manifestarse esta situación especial, la serie de tiempo tiende a poseer una dinámica semejante a través de los años presentando una tendencia al alza en los primeros

²⁸ UNGRD. Fenómeno el niño, Análisis comparativo 1997-1998//2014-2016. UNIDAD NACIONAL PARA LA GESTIÓN DEL RIESGO DE DESASTRES, 2016, págs. 1-143.

meses para después disminuir hasta alcanzar su menor valor a mediados del año.

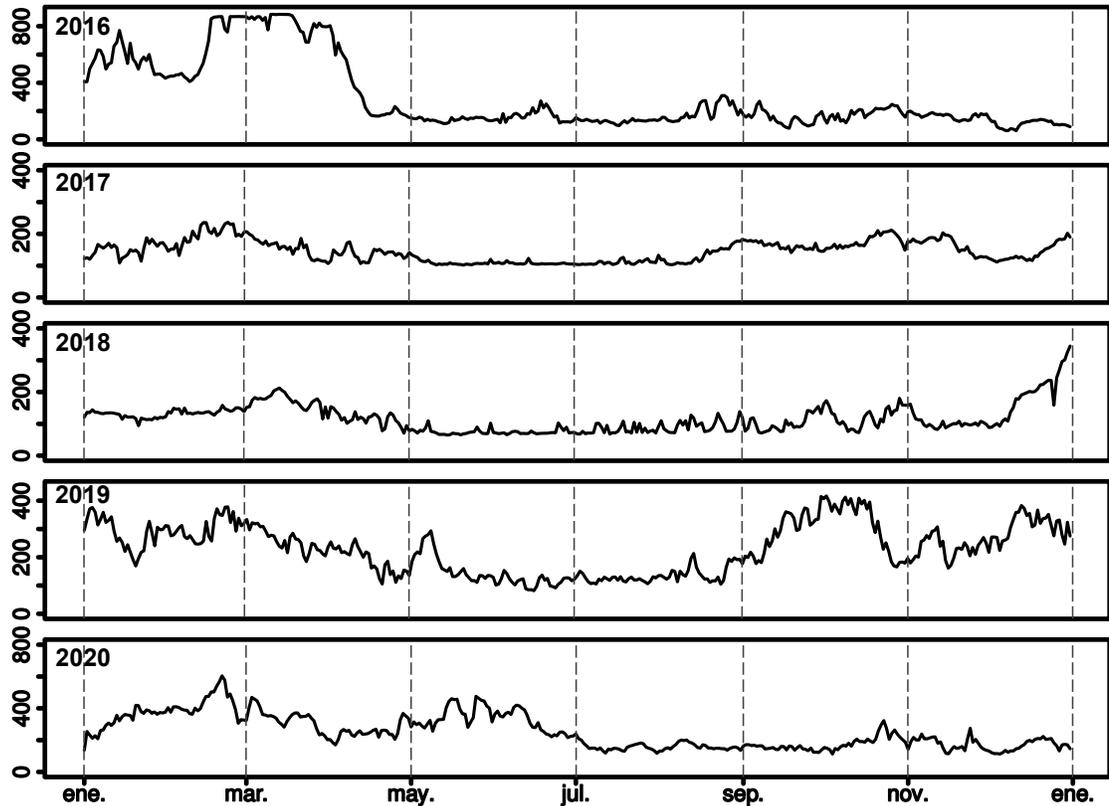


Figura 7. Descomposición de la serie de tiempo del precio de bolsa promedio del mercado de energía en Colombia por año.

Fuente: Autores.

De la misma manera, al comparar el comportamiento de las series, es posible reconocer que estas presentan una relación inversa debido a la presión al alza sobre el precio de bolsa que genera la escasez de los aportes hídricos. A saber, se observa que los valores más altos alcanzados por esta variable se manifestaron en los momentos en que el país poseía pocas cantidades de agua.

Teniendo en cuenta que las estructuras de las series de tiempo estudiadas presentan una dinámica que tiende a repetirse a través de los años, en este trabajo de investigación se seleccionó una ventana de tiempo deslizante de un intervalo de un año con la finalidad de llevar a cabo el entrenamiento de los modelos para la

elaboración de los pronósticos.

3.2. Implementación del modelo ARIMA

En el desarrollo de esta metodología se emplean únicamente los valores de la serie de tiempo de frecuencia diaria calculada para el precio de bolsa. A partir de una ventana de entrenamiento deslizante de un año se determina de manera automática por medio de la función `auto.arima` el modelo que mejor se adapta a la dinámica de la serie a fin de realizar el pronóstico del día siguiente para el último dato seleccionado. Es decir que, si se desea pronosticar el valor del precio de bolsa del día 01 de enero del año 2021 se requieren los datos comprendidos entre el 01 de enero y el 31 de diciembre del año 2020 para la estimación automática del modelo.

La función `auto.arima` de R utiliza un procedimiento iterativo que combina pruebas de raíces unitarias, minimización de criterios de información y estimación de máxima verosimilitud para seleccionar el mejor ajuste de los parámetros que conforman el modelo. El proceso iterativo ejecutado por el algoritmo se describe a continuación²⁹:

1. De manera inicial a través de la aplicación de la prueba de hipótesis de **Kwiatkowski-Phillips-Schmidt-Shin (KPSS)**, se determina el número de diferencias que se deben aplicar a la serie de tiempo para asegurar que esta constituya un proceso estacionario. En este punto de manera predeterminada el orden de las diferencias estará acotado entre los valores $0 \leq d \leq 2$.
2. Después de identificar el número de diferencias del modelo, utilizando los datos de la serie diferenciada se calcula la mejor estimación del orden de **p** y **q** minimizando el criterio de información de **Akaike (AIC)**. Para esta tarea, el algoritmo en lugar de considerar todas las combinaciones posibles, utiliza

²⁹ HYNDMAN. Op. Cit.

una búsqueda iterativa por pasos para recorrer todo el espacio. Los pasos que sigue la búsqueda se listan a continuación³⁰:

- a) De manera preliminar se eligen los siguientes modelos a fin de identificar por medio del criterio AIC el mejor de todos ellos.

ARIMA(0,d,0) ; ARIMA(2,d,2) ; ARIMA(1,d,0) ; ARIMA (0,d,1)

Se debe incluir una constante a menos de que el índice **d** sea igual a dos. Del mismo modo si el índice **d** es menor o igual a uno, se incluye un modelo adicional *ARIMA(0,d,0)* al conjunto anterior.

- b) El mejor modelo seleccionado en el paso anterior se define como el “modelo actual” del proceso.
- c) A partir del modelo actual se realizan variaciones de los parámetros **p** y/o **q** en ± 1 , y se incluye o excluye la constante agregada al modelo. En este paso el mejor modelo de todos los considerados en el proceso se designa como el nuevo modelo actual.
- d) Finalmente, se repite el paso anterior hasta garantizar el modelo con el valor de criterio de información más bajo.

El comando de R utilizado en este documento para la estimación automática del mejor modelo por medio de la función es el siguiente:

auto.arima(Pb_st) (10)

Donde **Pb_st** corresponde a una variable de R de tipo serie de tiempo que contiene los datos del precio de bolsa promedio en el intervalo analizado.

³⁰ Ibid.

3.3. Implementación del modelo de regresión dinámica

En la aplicación de esta técnica se utiliza la información de las dos series de tiempo con la intención de modelar el valor promedio del precio de bolsa en función de la disponibilidad y el comportamiento de los aportes hídricos. Al igual que con el modelo anterior se toma una ventana de entramiento deslizante de un año para seleccionar de manera automática mediante la ejecución de `auto.arima` el modelo que mejor se ajusta al patrón de los datos. La función `auto.arima` de R permite ajustar automáticamente un modelo de regresión lineal con errores ARIMA al suministrar en su argumento un vector con la información de la variable predictora³¹.

En la operación del mercado de energía eléctrica de Colombia el comportamiento de la variable del precio de bolsa se encuentra influenciado por las variaciones presentadas en días anteriores en los aportes hídricos, ya que por el mecanismo de negociación los agentes generadores deben suministrar con un día de anticipación la información de su precio de oferta para que el operador pueda desarrollar la programación del despacho³². Por esta razón, se debe modificar la ecuación del modelo de regresión para incluir la influencia de los efectos retardados de la variable predictora. Teniendo en cuenta que en este estudio se realizó la selección de un solo predictor (**aportes hídricos**), la ecuación del modelo que incluye los efectos retardados es la siguiente³³:

$$y_t = \beta_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \dots + \gamma_k x_{t-k} + \eta_t \quad (11)$$

³¹ Ibid.

³² RINCÓN SARABIA, Rolando Andrés y PABA ÁLVAREZ, Alberth Andrés. Análisis y modelado del precio de la energía eléctrica para la administración de recursos energéticos locales. Universidad Industrial de Santander, 2016, págs. 1-135.

³³ HYNDMAN. Op. Cit.

Con el propósito de identificar el número de retardos ("Lags" por su traducción al inglés) óptimos que se deben incluir en la estimación del modelo, se construyen con los datos del año 2020 diez modelos tentativos que contienen la información actual de la serie y otras versiones con hasta nueve días de retraso, a fin de verificar a través del criterio de información **AICc** el mejor de todos estos. Los resultados del criterio de información para cada modelo son los siguientes:

	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4
AICc	3360.342	3354.850	3350.377	3346.592	3341.709

	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9
AICc	3343.502	3343.855	3345.908	3346.693	3348.830

Tabla 1. Criterios AICc de los modelos de regresión tentativos.

Fuente: Autores.

Al comparar la magnitud de los criterios de información se observa que el modelo desarrollado con la serie independiente retrasada en cuatro unidades presenta el menor valor, indicando que una modificación en la variable de los aportes hídricos se demora aproximadamente cuatro días en influir el valor promedio del precio de bolsa. Por consiguiente, con base en este resultado se emplea un vector de la variable predictora con cuatro días de retardo para llevar a cabo la estimación automática del modelo. El comando de R ejecutado para esta labor es el siguiente:

$$auto.arima(Pb_st, xreg = \log(Ap_lagg)) \quad (12)$$

Donde **Pb_st** corresponde a una variable de R de tipo serie de tiempo que contiene los datos del precio de bolsa promedio en el intervalo analizado y **Ap_lagg** un vector de R con los datos de la serie de los aportes retrasados cuatro días.

4. Elaboración de pronósticos

El pronóstico a corto plazo del valor promedio del precio de bolsa después de realizar la estimación automática del mejor ajuste de los modelos se efectúa mediante la ejecución de la función **forecast** de R. En el caso de la técnica de regresión dinámica se utiliza la función `auto.arima` con la serie de los aportes hídricos para determinar el valor futuro de los predictores.

Con la intención de evaluar la funcionalidad y el rendimiento de la metodología planteada se seleccionó la ventana de tiempo comprendida entre el 01 de Julio y el 31 de diciembre del año 2019 a fin llevar a cabo el pronóstico del precio de bolsa.

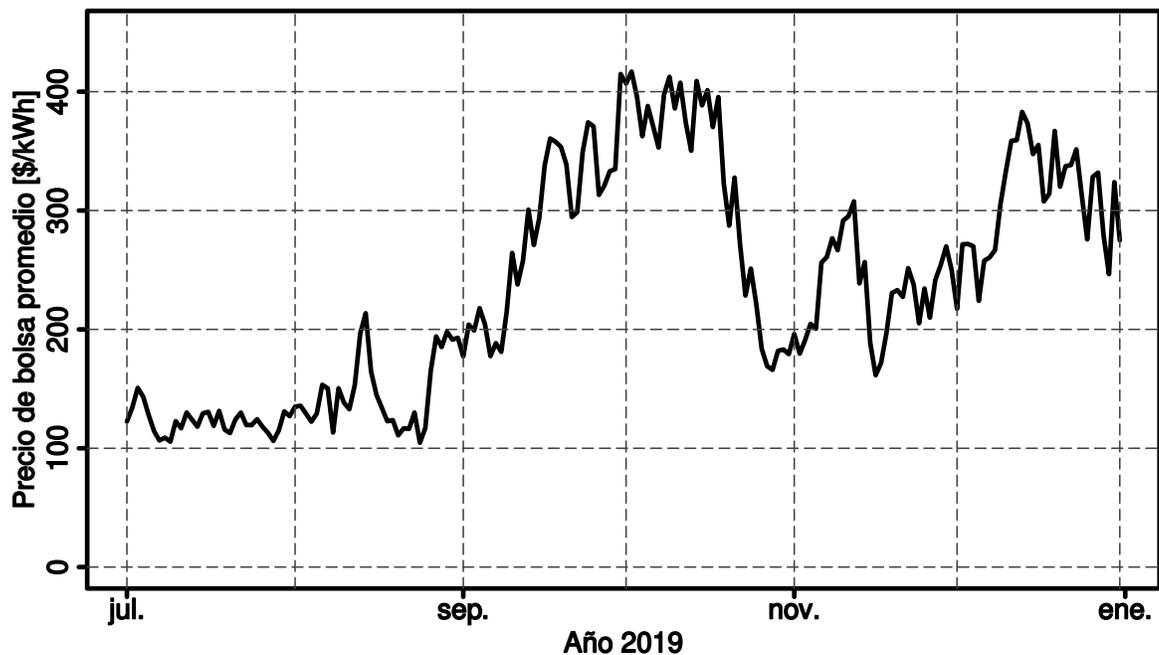


Figura 8. Serie de tiempo del precio bolsa promedio - Ventana seleccionada para la evaluación de los modelos.

Fuente: Autores.

La **Figura 8** muestra la variable del precio de bolsa en la ventana especificada.

Se realizó la elección de este intervalo de evaluación dado que en ese tiempo se presentó un comportamiento problemático para los modelos de predicción debido a la variación constante de la dinámica de la serie.

Para comparar la precisión y el desempeño de los pronósticos desarrollados con la metodología ARIMA y el método de regresión dinámica, se utilizaron las siguientes medidas de error:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (13)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} * 100 \quad (14)$$

Donde y_i es el valor real de la serie y \hat{y}_i el pronosticado.

Como se mencionó en la sección anterior, utilizando una ventana de entrenamiento deslizante de un año, iniciando con los datos comprendidos entre el 01 de julio del año 2018 y el 30 de junio del año 2019, se estimaron las predicciones del valor promedio del precio de bolsa modificando el horizonte de pronóstico hasta siete días, obteniendo los resultados mostrados en las **Figuras 9, 10 y 11**.

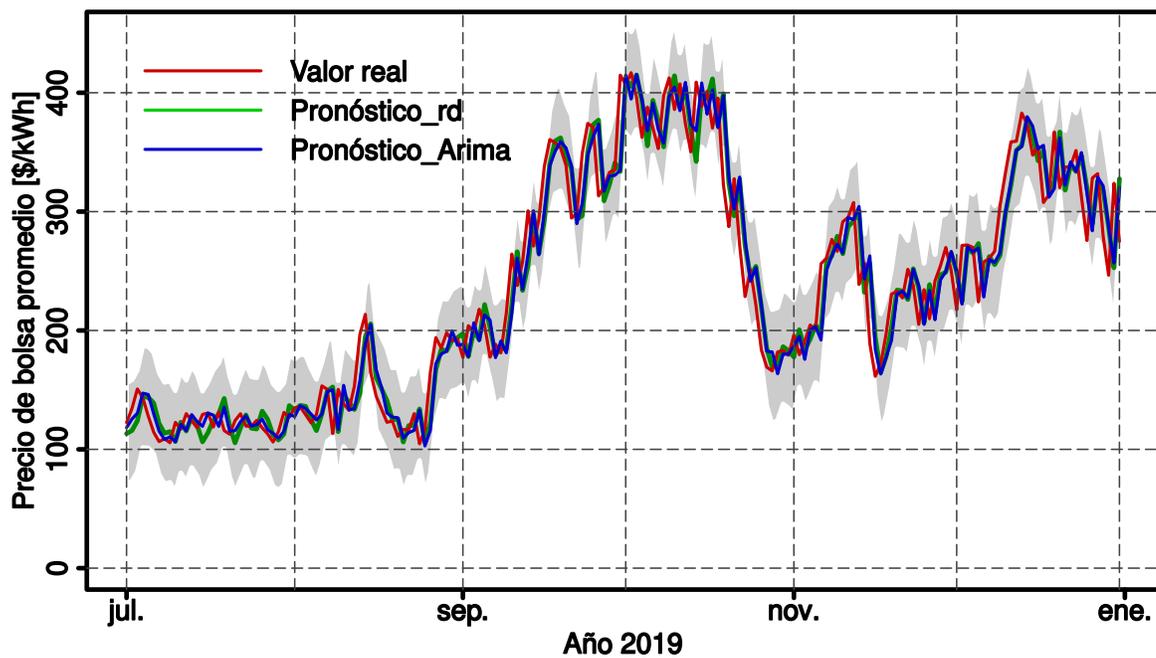


Figura 9. Pronóstico del precio de bolsa promedio con ventana de predicción de un día.

Fuente: Autores.

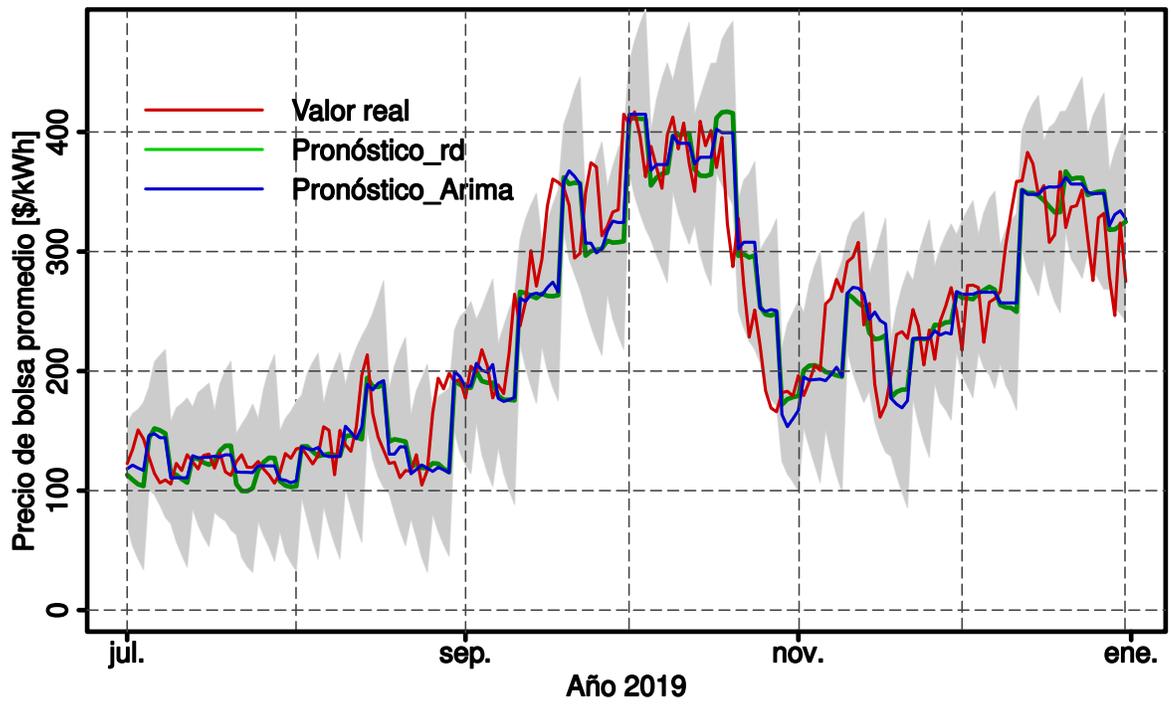


Figura 10. Pronóstico del precio de bolsa promedio con ventana de predicción de cuatro días.

Fuente: Autores.

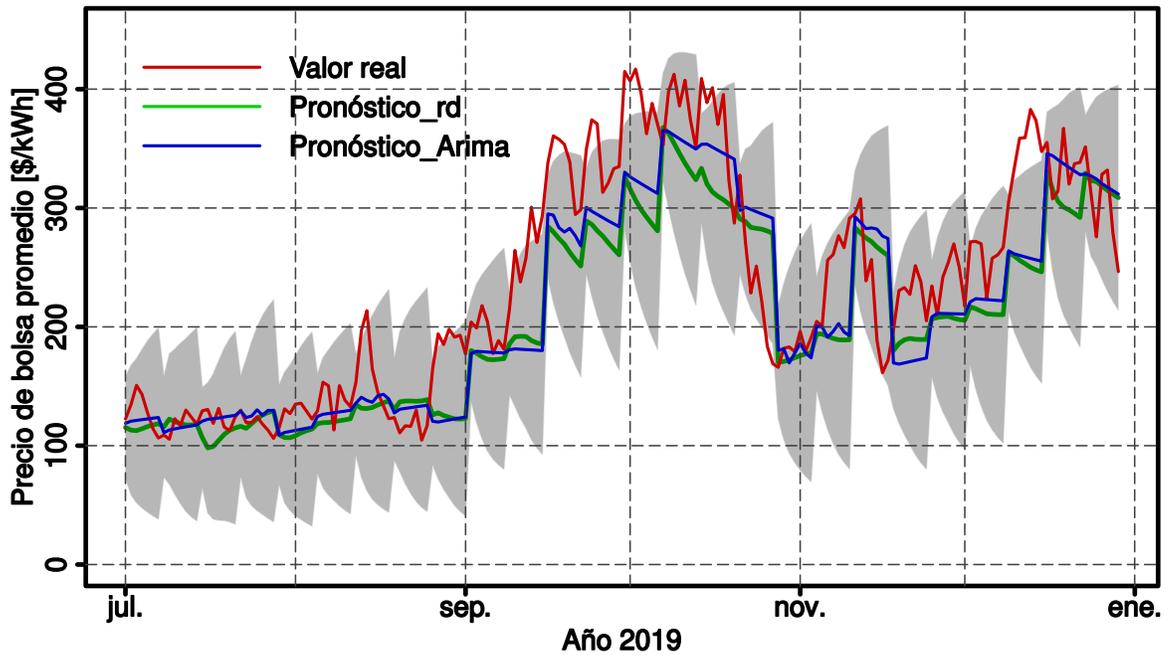


Figura 11. Pronóstico del precio de bolsa promedio con ventana de predicción de siete días.

Fuente: Autores.

Al examinar la dinámica de las curvas obtenidas en la elaboración de los pronósticos, se identifica que la **Figura 9** posee una buena concordancia entre los valores de la serie real y los valores pronosticados, debido a que los errores de predicción mínimos y máximos cometidos por cada modelo para este horizonte de pronóstico se encuentran dentro de límites aceptables, tal como se muestra en las **Tablas 2 y 3**.

Horizonte de pronóstico	MAE		MAPE	
	[COP/kWh]		[%]	
	MIN	MAX	MIN	MAX
Un día sin desfase	0.0000	17.5336	0.0000	7.6013
Un día	0.0663	80.0134	0.0508	39.1167
Dos días	0.5719	106.6220	0.2219	38.4808
Tres días	0.0663	119.9920	0.0508	63.5170
Cuatro días	0.3546	112.1043	0.1021	51.0173
Cinco días	0.0663	165.4983	0.0508	89.3916
Seis días	0.2202	127.5839	0.0650	69.1705
Siete días	0.1621	138.7614	0.3115	77.5273

Tabla 2. Error máximo y mínimo cometido en el pronóstico con la técnica ARIMA.
Fuente: Autores.

Horizonte de pronóstico	MAE		MAPE	
	[COP/kWh]		[%]	
	MIN	MAX	MIN	MAX
Un día sin desfase	0.0361	17.1613	0.0098	11.9033
Un día	0.3517	81.1028	0.1120	35.7672
Dos días	0.3517	104.7501	0.1120	36.9178
Tres días	0.2092	105.6183	0.1619	55.9084
Cuatro días	0.2677	128.4624	0.1308	50.1092
Cinco días	0.1278	161.6362	0.0491	83.6200
Seis días	0.4217	140.9908	0.2204	72.1340
Siete días	0.5704	131.0728	0.3115	77.5273

Tabla 3. Error máximo y mínimo cometido en el pronóstico con la técnica de regresión dinámica.
Fuente: Autores.

Por otra parte, en las **Figuras 10 y 11** se evidencia que a medida que aumenta el

intervalo de predicción se incrementa el error de los pronósticos. Esto porque los modelos estimados no pueden reaccionar de manera inmediata a las variaciones presentadas por la serie.

Realizando una comparación entre los datos estimados a partir de los modelos de predicción y los datos reales de la serie, se puede notar que en las tres figuras se manifiestan dos tipos de errores en el pronóstico: un error asociado a la magnitud y otro a la fase. Los errores de magnitud se encuentran relacionados con la capacidad de predicción de los métodos estudiados, mientras que los errores de fase son un compartimiento esperado cuando se efectúan pronósticos con la información de series de tiempo ³⁴. Los valores del error medio absoluto (**MAE**) y el error porcentual medio absoluto (**MAPE**) de los pronósticos elaborados se representan en las **Figuras 12 y 13**.

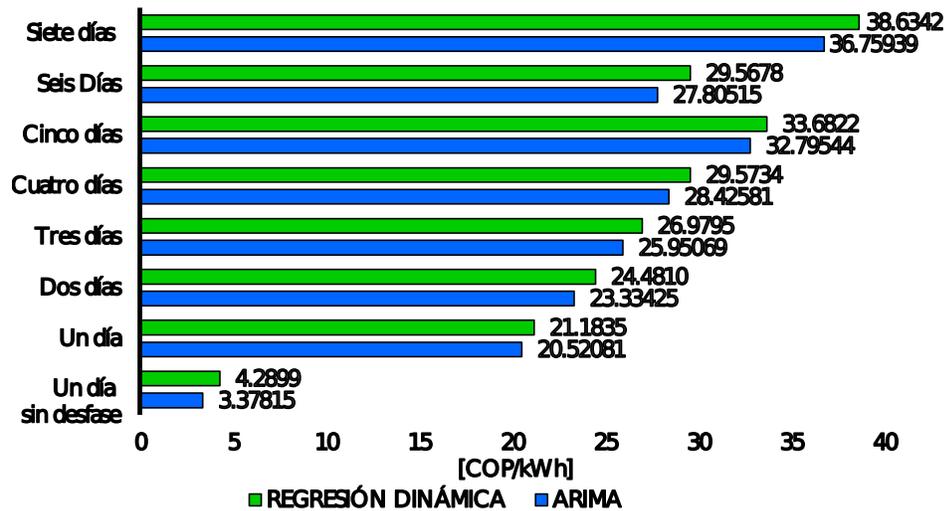


Figura 12. Valores de error medio absoluto de cada pronóstico.

Fuente: Autores.

³⁴ LIU, Xiaolei; LIN, Zi y FENG, Ziming. Short-term offshore wind speed forecast by seasonal ARIMA - A comparison against GRU and LSTM. En: Energy 227 (2021); págs. 1-12.

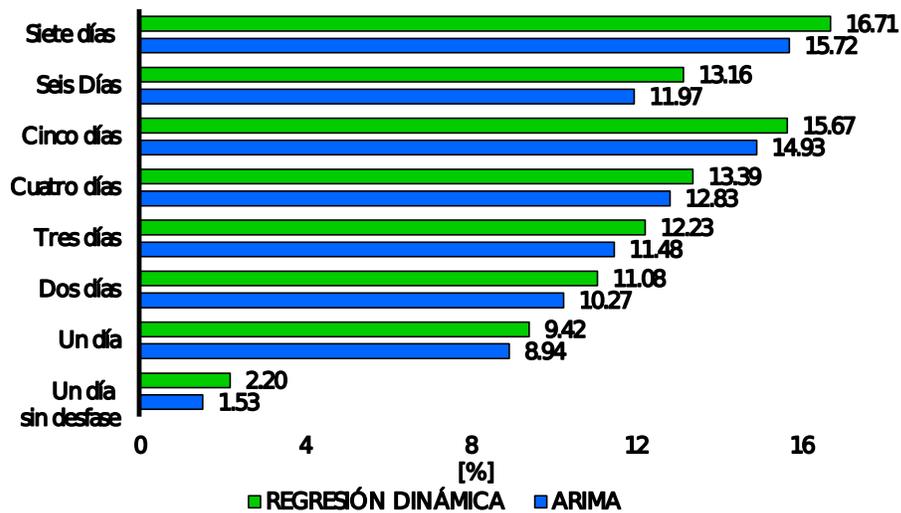


Figura 13. Valores de error porcentual medio absoluto de cada pronóstico.
Fuente: Autores.

De las **Figuras 12 y 13** se puede destacar que el menor termino de error de los pronósticos ejecutados por cada modelo se cometió cuando se llevó a cabo la predicción de los 184 valores del precio de bolsa para el año 2019 a partir de una ventana de pronóstico de un día, y eliminando el error de fase. También en función de estos resultados se deduce que el método calculado mediante la metodología ARIMA es ligeramente superior a la técnica de regresión dinámica.

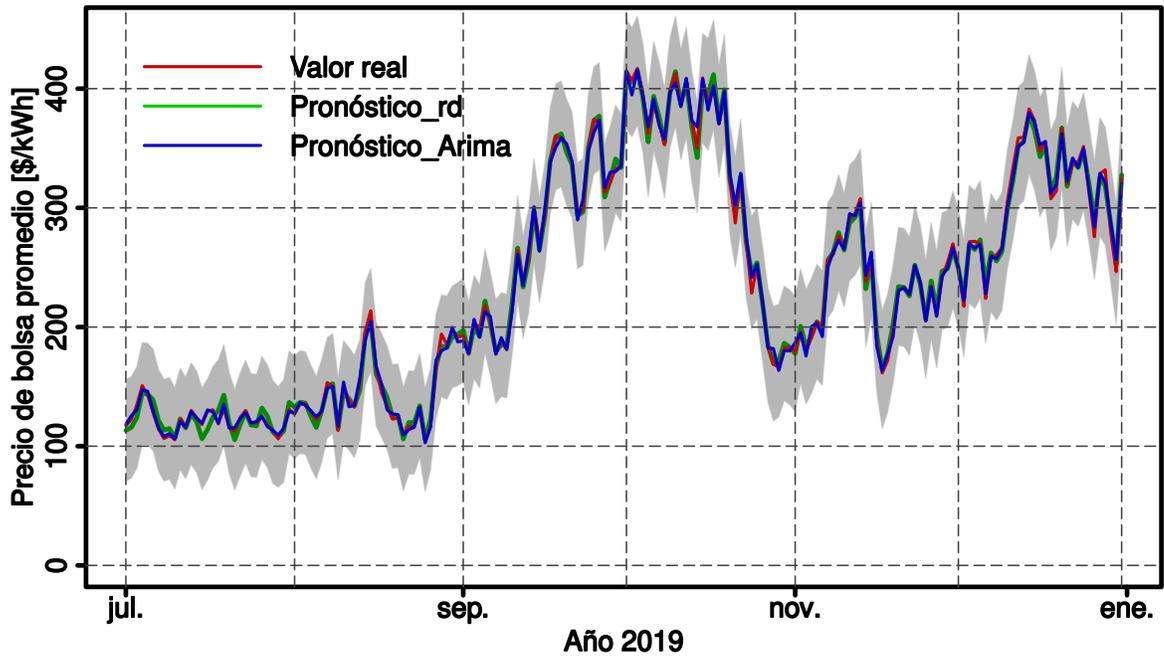


Figura 14. Pronóstico del precio de bolsa promedio con ventana de predicción de un día eliminando la componente de desfase.
Fuente: Autores.

5. Conclusión

Se desarrollaron dos metodologías para efectuar la predicción a corto plazo del precio de bolsa promedio del mercado de energía eléctrica de Colombia en función del análisis de la información de las series de tiempo. De los resultados se observa que, pese a los errores, ambos modelos probaron ser confiables en la elaboración de los pronósticos, puesto que lograron a través de su valor promedio replicar la dinámica real de la serie con porcentajes de error aceptables, demostrando que la variable de los aportes hídricos representa un predictor de importancia para el modelado del precio de bolsa del mercado colombiano. De las dos metodologías estudiadas, el enfoque de los modelos ARIMA comprobó ser el mejor debido a que logró superar la técnica de regresión dinámica sin tener en cuenta la influencia de la serie de tiempo de los aportes hídricos. En los dos escenarios se identificó que la estimación del valor futuro del precio de bolsa se vuelve mas imprecisa a medida que se aumenta el horizonte de pronóstico.

Como trabajo a futuro se propone estudiar una metodología que permita enlazar la información de las dos series de tiempo por medio de la técnica ARIMA a fin de mejorar la precisión de los pronósticos en el corto plazo disminuyendo la magnitud de los errores y eliminando la componente de error de fase introducida por la utilización de las series de tiempo.

Bibliografía

Bello Rodríguez, Sandra Patricia y Robert Baudilio Beltrán-Ahumada. “Caracterización y pronóstico del precio spot de la energía eléctrica en Colombia”. En: *Revista de la Maestría en Derecho Económico* 6.6 (2013), págs. 293-316.

Bissing, Daniel y col. “A Hybrid Regression Model for Day-Ahead Energy Price Forecasting”. En: *IEEE Access* 7 (2019), págs. 36833-36842. DOI: 10.1109/ACCESS.2019.2904432.

Botero-Botero, Sergio y Jovan Alfonso Cano-Cano. “Análisis de series de tiempo para la predicción de los precios de la energía en la bolsa de Colombia”. En: *Cuadernos de Economía* 27.48 (2008).

Contreras, J. y col. “ARIMA models to predict next-day electricity prices”. En: *IEEE Transactions on Power Systems* 18.3 (2003), págs. 1014-1020. DOI: 10.1109/TPWRS.2002.804943.

Fernández, Victoria. *ARIMA*. [En línea] Disponible en: <https://rpubs.com/vicferu/394457>. 2018.

Gao, Gao, Kwoklun Lo y Fulin Fan. “Comparison of ARIMA and ANN models used in electricity price forecasting for power market”. En: *Energy and Power Engineering* 09 (2017), págs. 120-126. DOI: 10.4236/epe.2017.94b015.

Hanke, John E. y Dean W. Wichern. *Pronosticos en los negocios*. Ed. por Antonio P Enriquez-Brito, Ofelia Vizcaíno-Díaz y José Antonio Rosas-Ferrer. Novena. Pearson Educacion, 2010, págs. 1-576.

Huertas, Nicolás y Santiago Donado. *Evaluación crítica de la función auto.arima de R en la identificación de procesos de series de tiempo*. Universidad de los Andes, 2018, págs. 1-33.

Hyndman, Rob J y George Athanasopoulos. *Forecasting: principles and practice*. Segunda. [En línea] Disponible en:
<https://otexts.com/fpp2/>. OTexts:Melbourne, Australia, 2018.

Jiang, LianLian y Guoqiang Hu. "A Review on Short-Term Electricity Price Forecasting Techniques for Energy Markets". En: *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. 2018, págs. 937-944. DOI: 10.1109/ICARCV.2018.8581312.

Jiménez-Navia, Breyner y Carlos Adolfo Moreno-Buitrago. *Modelos estadísticos tipo ARIMA para el pronóstico de incautaciones de drogas ilegales por parte de la armada nacional de Colombia*. Fundación universitaria los libertadores, 2017.

Liu, Xiaolei, Zi Lin y Ziming Feng. "Short-term offshore wind speed forecast by seasonal ARIMA - A comparison against GRU and LSTM". En: *Energy* 227 (2021), págs. 1-12. DOI: <https://doi.org/10.1016/j.energy.2021.120492>.

McHugh, Catherine y col. "Forecasting Day-ahead Electricity Prices with A SARIMAX Model". En: *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2019, págs. 1523-1529. DOI: 10.1109/SSCI44817.2019.9002930.

Montero-Granados, Roberto. *Modelos de regresión lineal múltiple*. Universidad de Granada, 2016, págs. 1-61.

Muñoz-Santiago, Alberto y col. “Pronóstico del precio de la energía en Colombia utilizando ARIMA con IGARCH”. En: *Revista de Economía del Rosario* 20 (2017), págs. 127-161. DOI: 10.12804/revistas.urosario.edu.co/economia/a.6152.

Rincón-Saravia, Rolando Andrés y Alberth Andrés Paba-Álvarez. *Análisis y modelado del precio de la energía eléctrica para la administración de recursos energéticos locales*. Universidad Industrial de Santander, 2016, págs. 1-135.

UNGRD. *Fenomeno el niño, Análisis comparativo 1997-1998//2014-2016*. UNIDAD NACIONAL PARA LA GESTIÓN DEL RIESGO DE DESASTRES, 2016, págs. 1-143.

XM. *PARATEC - Parámetros técnicos del SIN*. [En línea] Disponible en: <http://paratec.xm.com.co/paratec/SitePages/Default.aspx>. 2020.