

PREDICCIÓN DE LOS PRECIOS PROMEDIOS MENSUALES DE CONTRATOS DESPACHADOS EN EL MERCADO MAYORISTA DE ELECTRICIDAD EN COLOMBIA USANDO MÁQUINAS DE VECTORES DE SOPORTE*

Juan David Velásquez H. **

Carlos Jaime Franco C. ***

Yris Olaya M. ****

* Artículo de investigación científica y tecnológica. Es producto de la investigación realizada por los grupos de Mercados Energéticos y Computación Aplicada en el Modelado y la Predicción de Variables Económicas en Mercados de Energía. Patrocinado por la Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. El artículo se recibió el 24-09-2009 y se aprobó el 09-03-2010.

** Doctor en Ingeniería, Área de Sistemas Energéticos, Universidad Nacional de Colombia, Medellín, Colombia, 2009; Magister en Ingeniería de Sistemas, Universidad Nacional de Colombia, Medellín, 1997. Profesor asociado, Escuela de Sistemas, Facultad de Minas, Universidad Nacional de Colombia. Director del grupo de investigación Computación Aplicada. Miembro del grupo de investigación Mercados Energéticos, Facultad de Minas, Universidad Nacional de Colombia. Correo electrónico: jdvelasq@unal.edu.co.

*** Doctor en Ingeniería, Área de Sistemas Energéticos, Universidad Nacional de Colombia, Medellín, Colombia, 2002; Magister en Aprovechamiento de Recursos Hidráulicos, Universidad Nacional de Colombia, Medellín, 1996. Profesor asociado, Escuela de Sistemas, Facultad de Minas, Universidad Nacional de Colombia. Miembro de los grupos de investigación Mercados Energéticos y de Computación Aplicada, Facultad de Minas, Universidad Nacional de Colombia. Correo electrónico: cjfranco@unal.edu.co.

**** Doctora en Economía de los Recursos Minerales, Colorado School of Mines, Colorado, Estados Unidos, 2006; Magister en Ingeniería de Sistemas, Universidad Nacional de Colombia, Medellín, Colombia, 1999. Profesora asociada, Escuela de Sistemas, Facultad de Minas, Universidad Nacional de Colombia. Directora del grupo de Mercados Energéticos. Miembro del grupo de Computación Aplicada, Facultad de Minas, Universidad Nacional de Colombia. Correo electrónico: yolayam@unal.edu.co.

Predicción de los precios promedios mensuales de contratos despachados en el mercado mayorista de electricidad en Colombia usando máquinas de vectores de soporte

RESUMEN

La predicción de los precios de la electricidad en los mercados liberalizados y desregulados ha sido considerada una tarea difícil, debido a la cantidad y complejidad de factores que gobiernan los precios. En este artículo se pronostican los precios promedios mensuales de los contratos despachados en el mercado eléctrico de Colombia usando una novedosa red neuronal, conocida como máquina de vectores de soporte. Se comparan los pronósticos obtenidos con un perceptrón multicapa y un modelo ARIMA. Los resultados obtenidos muestran que la máquina de vectores de soporte captura de mejor forma la dinámica intrínseca de la serie de tiempo y es capaz de pronosticar con mayor precisión para un horizonte de 12 meses adelante.

Palabras clave:

estudios comparativos, series no lineales, predicción, precios de electricidad, redes neuronales.

Forecasting Average Monthly Prices for Colombian Wholesale Electricity Market Contracts

ABSTRACT

Forecasting electricity prices in liberalized, deregulated markets has always been considered a difficult task, due to the number of factors that govern prices and to their complexity. This article predicts the average monthly prices for Colombian electricity market contracts by using a novel neural network known as the support vector machine. Forecasts obtained using a multilayer perceptron are compared to forecasts obtained using an ARIMA model. The results show that the support vector machine better captures the intrinsic dynamics of the time series and is able to make more precise forecasts considering a 12-month horizon.

Key words:

Comparative studies, non-linear series, prediction, electricity prices, neural networks.

Predição dos preços médios mensais de contratos despachados no mercado atacadista de eletricidade na Colômbia usando máquinas de vetores de suporte

RESUMO

A predição dos preços da eletricidade nos mercados liberalizados e desregulados tem sido considerada uma tarefa difícil, devido à quantidade e complexidade de fatores que governam os preços. Neste artigo prognosticam-se os preços médios mensais dos contratos despachados no mercado elétrico da Colômbia usando uma inovadora rede neural, conhecida como máquina de vetores de suporte. Comparam-se os prognósticos obtidos com um perceptron multicamada e um modelo ARIMA. Os resultados obtidos mostram que a máquina de vetores de suporte captura de melhor forma a dinâmica intrínseca da série de tempo e é capaz de prognosticar com maior precisão para um horizonte de 12 meses adiante.

Palavras chave:

estudos comparativos, séries não lineares, predição, preços de eletricidade, redes neurais.

Introducción

Con la entrada en vigencia de la Ley de Servicios Públicos Domiciliarios (Ley 142) y la Ley Eléctrica (Ley 143), se dio lugar a la reestructuración del sector eléctrico colombiano. Los objetivos primordiales de esta reforma buscaban corregir los problemas que se presentaron en años anteriores, en especial en lo referente a los racionamientos ocurridos entre marzo de 1991 y mayo de 1992, que llevaron a la crisis al sector eléctrico. Así, el nuevo esquema de mercado promueve la libre competencia como el mecanismo primordial para forzar la eficiencia del mercado y garantizar la expansión requerida por el sistema. Bajo el nuevo esquema, la comercialización de electricidad en el sector se realiza a través de dos mecanismos fundamentales: los contratos bilaterales entre agentes y la subasta en la Bolsa de Energía, en la cual los distintos agentes reportan la cantidad de energía disponible y su respectivo precio para la venta, de tal forma que la Bolsa fija el precio de venta para los compradores diariamente.

La predicción de precios de bolsa es un problema especialmente difícil (Velásquez, Dyner y Souza, 2007), debido a la cantidad y complejidad de los factores que influyen en su determinación, como las características físicas del sistema de generación, la influencia de las decisiones de negocio de los distintos agentes y la regulación. En general, los precios de bolsa manifiestan dichas complejidades a través de sus características, que incluyen: pronunciados ciclos estacionales de periodicidad diaria, semanal, mensual y demás; volatilidad variable en el tiempo y regiones de volatilidad similar; fuertes varia-

ciones de año a año y de estación a estación, estructura dinámica de largo plazo, efectos de apalancamiento y respuesta asimétrica de la volatilidad a cambios positivos y negativos, valores extremos, correlaciones de alto orden, cambios estructurales, tendencias locales y reversión en la media, diferentes determinantes para los riesgos de corto, mediano y largo plazo, dependencia de las condiciones de las unidades de generación en el corto plazo y de las inversiones en capacidad y crecimiento de la demanda en el largo plazo.

Ante la complejidad de la dinámica de los precios de Bolsa, la dificultad de su pronóstico y su riesgo implícito, los contratos son un mecanismo de mitigación de riesgo que facilita la operación comercial de los diferentes agentes del mercado. Por una parte, evitan que el comprador se vea sujeto a la variabilidad de los precios en la Bolsa y a precios excepcionalmente altos, que ocurren ante la presencia de eventos hidrológicos extremos secos; por otra parte, estabilizan los ingresos del vendedor y lo protegen de precios excepcionalmente bajos, que ocurren cuando se presentan eventos hidrológicos extremadamente húmedos. Existen dos tipos de contratos representativos en el mercado eléctrico colombiano: pague-lo-contratado y pague-lo-demandado.

El tipo pague-lo-contratado especifica que el comprador se compromete a pagar toda la electricidad contratada, independientemente de si ella fue consumida o no; si se contrató una mayor cantidad de electricidad a la consumida, el excedente es vendido por el comprador a la Bolsa de Energía. En el tipo pague-lo-demandado, el comprador sólo paga

la energía efectivamente consumida; usualmente, se fija un límite superior a la energía que puede ser consumida por el comprador.

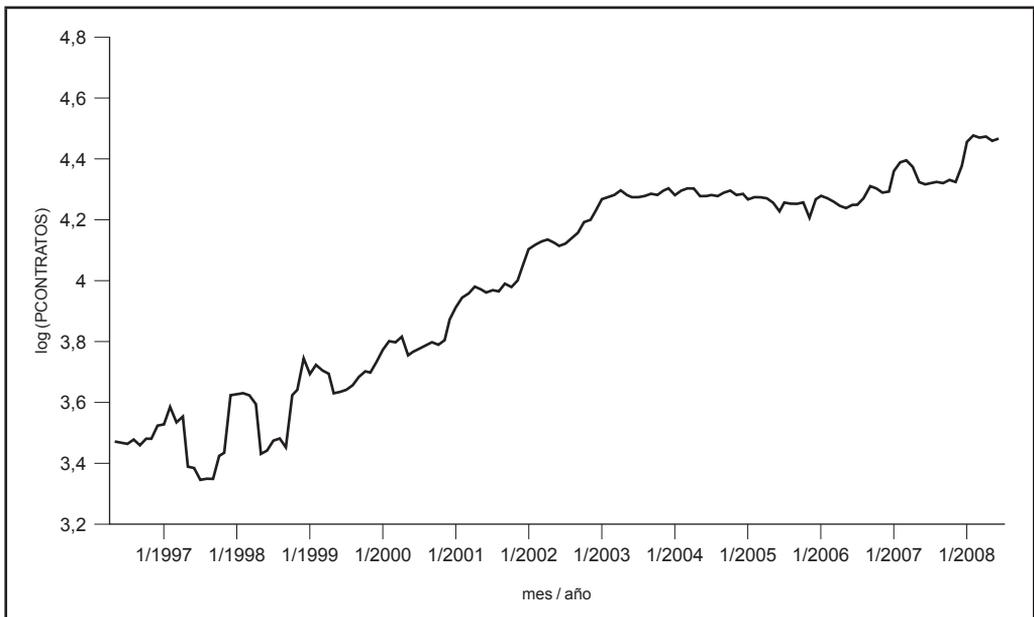
La predicción de los precios de la electricidad en el mercado, tanto en Bolsa como en contratos, se constituye en un insumo fundamental que requieren los diferentes agentes para establecer sus políticas de comercialización, donde se especifican las cantidades de energía vendida a través de cada uno de los mecanismos de negociación, de acuerdo con las expectativas de precios. Consecuentemente, se requiere desarrollar modelos de predicción para dichas series. Para el mercado colombiano, esta serie corresponde a los

precios promedios mensuales de los contratos despachados en la Bolsa de Energía.

Su traza en el tiempo es presentada en el Gráfico 1. Se intuye de antemano que su predicción es una tarea difícil, ya que hereda las complejidades propias del mercado eléctrico. La presencia de no linealidades en las series de precios de la electricidad ha motivado el uso de técnicas no lineales para su predicción; en el caso lineal, se parte de que el precio actual puede ser pronosticado a partir de los precios pasados como una combinación lineal de ellos:

Gráfico 1

Logaritmo natural del precio promedio mensual de los contratos despachados en el mercado colombiano



Fuente: Sistema Neón.

$$y_t = c + \sum_{p=1}^P \phi_p y_{t-p} + e_t \tag{1}$$

Donde y_t es la serie de precios, c es una constante y e_t es una variable aleatoria normal con media cero y varianza constante desconocida. Existen dos formas de no linealidad: en la primera se asume que la varianza de e_t cambia en el tiempo, mientras que en la segunda la relación existente entre el precio actual y_t y sus valores pasados no sigue (1). Para este último caso, al sustituir la sumatoria en (1) por otra forma funcional, se pueden obtener diferentes clases de modelos no lineales, como los modelos de transición brusca (Tong, 1978) o suave (Teräsvirta, 1994; Teräsvirta, Lin y Granger, 1993), los modelos autorregresivos de coeficientes funcionales (Chen y Tsay, 1993), diferentes tipos de redes neuronales artificiales y muchos otros. Una revisión detallada es presentada por Franses y van Djck (2000).

Las máquinas de vectores de soporte (SVM, por su sigla en inglés) (Vapnik, 1995; Vapnik, Golowich y Smola, 1996) son un tipo de red neuronal artificial desarrollada originalmente para solucionar problemas no lineales de clasificación (Belousov, Verzakov, y von Frese, 2002; Burges, 1998), pero que recientemente ha sido aplicada a la predicción de series temporales (Mukherjee, Osuna, y Girosi, 1997; Muller, Smola, Ratsch, Scholkopf, Kohlmorgen y Vapnik, 1997; Osowski y Garanty, 2007; Tay y Cao, 2001; Thissen, van Brakel, de Weijer, Melssen y Buydens, 2003).

Un resumen de los principales estudios seleccionados se presenta en el Cuadro 1. Su éxito,

se debe a que poseen una alta capacidad de generalización (Scholkopf y Smola, 2002; Vapnik, 1998), que es función directa de su estructura y de la metodología utilizada para estimar sus parámetros. Las SVM presentan varias ventajas: primero, su especificación se basa en el principio de minimización del riesgo estructural, que equivale a minimizar el límite superior del error de generalización del modelo. Así, la estimación de una SVM busca encontrar tanto la estructura óptima del modelo como los valores óptimos de sus parámetros, lo que permite una mayor capacidad de generalización.

Cuadro 1

Estudios seleccionados sobre la predicción de series de tiempo usando SVM

Estudio	Año	Modelos	Serie
Mukherjee et al.	1997	SVM Polinomios Redes RBF MLP	Mackey-Glass Mapa de Ikeda Lorenz
Muller et al.	1997	SVM RBF	Mackey-Glass Serie D-Santa Fe Competition
Tay y Cao	2001	SVM	5 series de precios en diferencias relativas
Thissen et al.	2003	SVM	Serie sintética generada con un ARIMA
		Red de Elman	Mackey-Glass
		MLP	Diferencias de presión en un filtro
Osowski y Garanty	2007	Wavelets + SVM	Polución atmosférica

Fuente: elaboración propia.

La estimación de los parámetros de otros modelos tradicionales, como los perceptro-

nes multicapa (MLP, por su sigla en inglés) (Zhang, Patuwo y Hu, 1998), se basa en la aplicación del principio de minimización de riesgo empírico, que depende fundamentalmente del ajuste a los datos históricos, tomando la estructura del modelo como un parámetro de entrada. Por lo tanto, dado que estos modelos son aproximadores universales de funciones, es común que se presente el fenómeno de sobreajuste, que consiste en que el MLP memoriza la información utilizada para la estimación de sus parámetros, pero los pronósticos son completamente lejanos de la realidad. Más aún, la especificación de los MLP se basa fundamentalmente en criterios heurísticos y el juicio experto del modelador (Masters, 1993 y 1995), de tal manera que dicho proceso se basa en un conjunto de pasos críticos que afectan el desempeño final del modelo (Kaastra y Boyd, 1996), en términos del ajuste a los datos históricos y su capacidad de generalización. Esto dificulta enormemente la reconstrucción del proceso de especificación y selección de esta clase de modelos.

La segunda ventaja de las SVM radica en que la estimación de los parámetros de una SVM es equivalente a la solución de un modelo de programación cuadrática con restricciones lineales. Ello implica que la solución óptima es global y única, lo cual representa una clara ventaja sobre modelos, como los MLP, que se caracterizan por poseer múltiples puntos de mínima local. No obstante, las SVM, al igual que otros tipos de modelos no paramétricos de regresión no lineal, como los MLP, sufren del problema del sobreajuste, lo que puede degradar ostensiblemente su capacidad de generalización.

El objetivo principal de este artículo es predecir los precios promedios mensuales despachados en la Bolsa de Energía del mercado eléctrico colombiano usando SVM, y comparar los resultados obtenidos con el modelo autorregresivo integrado de promedios móviles (ARIMA) y los MLP, con el fin de determinar el mejor modelo de predicción para la serie estudiada. La originalidad e importancia de la investigación propuesta está basada en los siguientes aspectos:

- Mientras que existe una extensa experiencia en la predicción de precios de electricidad en mercados de corto plazo (véase el trabajo de Velásquez, Dyner y Souza, 2007), no existen referencias en la literatura sobre la predicción de precios de contratos. En cuanto a este aspecto se da la primera contribución de este artículo, ya que se pretende desarrollar varios modelos de pronóstico para dicha serie.
- Existen pocas experiencias reportadas en la literatura que comparen el desempeño de las SVM con otros modelos al pronosticar series del mundo real. La serie analizada en esta investigación es particularmente difícil debido a la complejidad y cantidad de sus determinantes, por lo que resulta especialmente adecuada para este fin.
- Se contribuye a difundir el uso de las SVM para la predicción de series de precios en mercados de electricidad, y ello aumenta así la cantidad de herramientas disponibles.

En su orden, se describe la información y metodología utilizadas (sección 1), los resultados obtenidos (sección 2) y las principales conclusiones obtenidas.

1. Información y metodología

1.1 Información utilizada

Los datos con que se elaboró este estudio corresponden al logaritmo natural de la serie de precios promedios mensuales de los contratos despachados en la bolsa del mercado mayorista de electricidad en Colombia, expresada en \$/kWh, entre 1996:5 y 2008:6, los cuales están disponibles en el Sistema Neón. La serie presenta (Gráfico 1) una tendencia creciente de largo plazo desde 1996:5 hasta el primer semestre del 2003. Durante ese mismo intervalo de tiempo se evidencia una componente cíclica de periodicidad anual de amplitud variable, explicada, posiblemente, por el ciclo invierno-verano.

La mayor amplitud de la componente periódica coincide con el fenómeno de El Niño, ocurrido entre 1997 y 1998. Esta componente cíclica, aunque no con una amplitud tan marcada, permanece hasta principios del 2004. Desde el 2003 se presenta una tendencia ligeramente descendente que finaliza en algún momento del primer semestre del 2006. Se evidencia en este momento del tiempo un cambio estructural en la serie, tanto en su tendencia como en su componente cíclica: por una parte, se recuperan los niveles de crecimiento que caracterizaron los años 2000, 2001 y 2002; por la otra, se presenta nuevamente un ciclo estacional de período anual, cuyo nivel más alto coincide con la estación de verano.

1.2 Metodología empleada

Sea una serie de tiempo, y_t , con regresores x_t , para la cual se poseen D ejemplos representativos. Una SVM permite aproximar y_t a través de la función:

$$\hat{y}_t = b + \sum_{d=1}^D w_d \times k(x_t, x_d) \tag{2}$$

Donde b es una constante y w_d son los factores de ponderación de la función de núcleo $k(\cdot, \cdot)$. Así, una SVM es la combinación lineal del mapeo de x_t en un espacio de características altamente no lineales, definido por los puntos x_d y la función de transformación no lineal $k(\cdot, \cdot)$.

En (2), $k(x_t, x_d)$ representa una función de núcleo que permite llevar el punto x a un espacio altamente dimensional parametrizado por los puntos x_d . Existen varias funciones que son típicamente utilizadas como núcleos (Gunn, 1997; Tay y Cao, 2001; Thissen et al., 2003), pero las experiencias reportadas indican que la función gaussiana o RBF:

$$k(x_t, x_d) = \exp\left(-\frac{\|x - x_d\|^2}{\sigma^2}\right) \tag{3}$$

produce mejores resultados. El parámetro σ^2 corresponde a la varianza de la campana gaussiana y es dependiente del problema específico.

La estimación de (2) se basa en minimización de la función de riesgo regularizado, $R(C, \epsilon)$, definida como:

$$R(C, \epsilon) = C \frac{1}{D} \sum_{d=1}^D L_{\epsilon}(y_d, \hat{y}_d) + \frac{1}{2} \sum_{d=1}^D w_d^2 \tag{4}$$

Donde el primer sumando mide el error empírico entre el modelo y los datos, mientras que el segundo corresponde a la componente de regularización y depende únicamente de los pesos w_d . La constante de regularización C permite variar la importancia de cada una de las componentes; así, valores muy altos de C enfatizan el ajuste del modelo a los datos, sin que importe qué tan grandes deban ser los pesos w_d para conseguirlo. Sin embargo, se sabe que el modelo pierde su capacidad de generalización de los datos a medida que los pesos w_d aumentan en magnitud, ya que ellos suelen causar una varianza excesiva en el modelo. Cuando C tiende a cero, la magnitud de la función $R(C, \epsilon)$ depende únicamente de w_d , sin importar el ajuste a los datos, haciendo que los pesos w_d disminuyan tanto como sea posible. L_{ϵ} es la función de error ϵ -insensible de Vapnik (1998) definida como:

$$L_{\epsilon}(y_d, \hat{y}_d) = \begin{cases} |y_d - \hat{y}_d| - \epsilon & \text{para } |y_d - \hat{y}_d| > \epsilon \\ 0 & \text{en otro caso} \end{cases} \tag{5}$$

Donde la constante ϵ representa la precisión deseada, y es interpretada como el radio de un tubo dentro del cual el error se considera cero. Este comportamiento es esquematizado en el Gráfico 1, donde los puntos negros representan los datos disponibles, que están alejados a una distancia mayor de ϵ unidades de la predicción del modelo, y para los cuales el error se considera superior a cero. Los puntos blancos representan los datos dentro

del túnel de tolerancia del modelo, y para los cuales el error es cero. Mientras menor sea el valor dado a ϵ , mayor es la precisión exigida al modelo, ya que se tendrá en cuenta una mayor cantidad de puntos que contribuyen al error de ajuste del modelo.

La solución de (2) puede obtenerse mediante la teoría de multiplicadores de Lagrange, y es fácilmente demostrable que el problema original es equivalente a resolver el problema de programación cuadrática (Platt, 1998; Vapnik, 1998):

$$\min_z \frac{1}{2} z' H z + f' z \tag{6}$$

Con restricciones lineales:

$$\sum_{d=1}^D (a_d - a_d^*) = 0 \tag{7}$$

$$0 \leq a_d, a_d^* \leq C \tag{8}$$

Donde a_d y a_d^* son los multiplicadores de Lagrange asociados a los datos, los cuales cumplen con las restricciones $a_d \times a_d^* = 0$, $a_d \geq 0$ y $a_d^* \geq 0$; $z = [a_d, a_d^*]'$; $f = [\epsilon - y_p, \epsilon + y_i]$; y finalmente:

$$H = \begin{bmatrix} -K & K \\ K & -K \end{bmatrix} \tag{9}$$

En esta última ecuación, K es una matriz de orden $D \times D$, con $k_{ij} = k(x_i, x_j)$. Los pesos w_d se obtienen como:

$$w_d = a_d - a_d^* \tag{10}$$

En la práctica, esto equivale a que $|w_d| \leq C$. La formulación como un problema de programación cuadrática implica: (a) que la solución obtenida es global, ya que la formulación del problema es convexa; (b) que la solución es única, pues la función de costo es estrictamente convexa, y (c) que muchos de los pesos w_d serán cero. De esta forma, solamente una fracción de los datos originales contribuye en la función de error. Dicha conclusión se basa en las condiciones de Karush-Kuhn-Tucker para la programación cuadrática.

Los puntos para los cuales sus correspondientes multiplicadores de Lagrange son diferentes de cero (esto es, con $L_\varepsilon(\cdot, \cdot) > 0$) son llamados vectores de soporte, y son los que permiten realizar la aproximación de y_t en (1). Así, se espera que a medida que aumente el valor de ε , disminuya la cantidad de vectores de soporte; pero que también disminuya la capacidad de aproximación del modelo a los datos originales.

Un problema fundamental en la metodología es que el modelo obtenido es una función de los parámetros C , ε y σ , los cuales deben ser ajustados para cada problema específico.

2. Resultados obtenidos y discusión

Tal como ya se indicó, la muestra utilizada corresponde al logaritmo natural de la serie de precios mensuales promedio de los contratos despachados en la bolsa del mercado mayorista de electricidad en Colombia, expresada en \$/kWh, entre 1996:5 y 2008:6. La serie consta de 146 observaciones, de las cuales las primeras 122 se usan para estimar

los parámetros de todos los modelos, y las siguientes 24 para evaluar su capacidad de generalización.

Con el fin de determinar la capacidad de generalización para distintos horizontes de tiempo, se usaron dos muestras de pronóstico: la primera consta de 12 observaciones (entre 2006:7 y 2007:6), y la segunda, de 24 observaciones (entre 2006:7 y 2008:6). Los horizontes de predicción seleccionados son especialmente difíciles, ya que vuelve a aparecer un patrón cíclico anual que había desaparecido desde el 2003 y porque en el punto donde se inicia la muestra de predicción se presenta un cambio de tendencia en la serie.

La bondad del ajuste fue medida mediante la sumatoria del error cuadrático medio (SSE, por su sigla en inglés) de los residuales, e_t :

$$SSE = \sum_{t=1}^R e_t^2 \tag{11}$$

Y su desviación media absoluta (MAD, por su sigla en inglés):

$$MAD = \frac{1}{R} \sum_{t=1}^R |e_t| \tag{12}$$

Se estimaron SVM que difieren en los rezagos considerados para pronosticar la serie. Para cada uno de los modelos se realizó la afinación manual de los parámetros C , ε y σ . Así, en el Cuadro 2, para cada conjunto de rezagos considerado, se reporta la mejor combinación encontrada para los parámetros C , ε y σ . En este mismo cuadro se reportan el SSE y el MAD para la muestra de calibración y para las dos muestras de predicción usadas

para validar la generalización de cada configuración considerada.

El modelo SVM-13, que usa como entradas los rezagos 1 al 13 de la serie, presenta el mejor ajuste a la muestra de calibración de los parámetros. Le sigue en calidad el modelo SVM-17, con un SSE superior en más de 300% al obtenido para el SVM-13. Los modelos SVM-2 y SVM-15 tienen el menor SSE

para la predicción con horizonte de un año; el SVM-2 presenta el peor ajuste a la muestra de calibración, por lo que el estadístico obtenido para la muestra de predicción puede considerarse coincidental.

Para el horizonte de predicción de 24 meses, el mejor ajuste es obtenido con el modelo SVM-15; para este, la diferencia entre los SSE calculados para los dos horizontes de predicción

Cuadro 2

Sumatoria del error cuadrático medio (SSE) y desviación media absoluta (MAD): SVM con Kernel gaussiano

Modelo	Rezagos	C	ϵ	σ	Entrenamiento SSE (MAD)	Predicción SSE (MAD) 1 año	Predicción SSE (MAD) 2 años
SVM-1	1	16,2	0,03	0,5	0,183 (0,023)	0,018 (0,029)	0,104 (0,050)
SVM-2	1, 2	12,1	0,01	0,6	0,186 (0,022)	0,012 (0,024)	0,080 (0,043)
SVM-3	1-3	12,1	0,01	0,6	0,161 (0,022)	0,013 (0,025)	0,110 (0,050)
SVM-4	1-4	12,1	0,01	0,7	0,152 (0,021)	0,013 (0,025)	0,110 (0,051)
SVM-5	1-5	12,1	0,01	0,7	0,130 (0,020)	0,016 (0,028)	0,146 (0,059)
SVM-6	1-6	11,0	0,02	0,7	0,103 (0,020)	0,015 (0,027)	0,122 (0,056)
SVM-7	1-7	11,0	0,02	0,7	0,093 (0,019)	0,015 (0,026)	0,128 (0,057)
SVM-8	1-8	11,0	0,02	0,8	0,067 (0,017)	0,014 (0,026)	0,121 (0,056)
SVM-9	1-9	11,0	0,02	0,8	0,062 (0,017)	0,015 (0,027)	0,121 (0,056)
SVM-10	1-10	11,0	0,02	0,9	0,057 (0,017)	0,014 (0,026)	0,101 (0,052)
SVM-11	1-11	11,0	0,02	0,9	0,048 (0,016)	0,017 (0,028)	0,133 (0,059)
SVM-12	1-12	3,2	0,02	1,0	0,067 (0,018)	0,027 (0,037)	0,240 (0,079)
SVM-13	1-13	3,8	0,01	0,5	0,014 (0,009)	0,024 (0,037)	0,210 (0,071)
SVM-14	1, 12	8,1	0,02	0,5	0,152 (0,022)	0,015 (0,026)	0,112 (0,051)
SVM-15	1, 12, 13	8,1	0,02	0,5	0,076 (0,018)	0,012 (0,023)	0,052 (0,035)
SVM-16	1, 5, 12, 13	8,7	0,02	0,6	0,050 (0,016)	0,016 (0,028)	0,106 (0,051)
SVM-17	1, 5, 8, 12, 13	8,7	0,02	0,6	0,047 (0,015)	0,014 (0,025)	0,105 (0,050)

Fuente: elaboración propia.

muestran que el segundo año es particularmente difícil, ya que el error obtenido respecto al primer año se triplica. Teniendo en cuenta simultáneamente las SSE para la muestra de calibración y un horizonte de predicción de un año se obtiene el modelo SVM-13, que tiene el mejor ajuste (con un SSE de 0,038), seguido del SVM-17 con un SSE de 0,061.

Para evaluar la capacidad de predicción de las SVM respecto a otros modelos, se consideraron como modelos competidores un MLP y un modelo ARIMA. Ambos modelos fueron estimados para diferentes conjuntos de rezagos y se seleccionaron los mejores modelos a partir de su calidad a las distintas muestras de datos.

La arquitectura del MLP consta de una capa de entrada con una neurona por cada uno de los rezagos considerados; una capa oculta con 1, 2 o 3 neuronas, y una capa de salida. La capa oculta se activa con la función sigmoidea bipolar; mientras que la capa de salida usa la función identidad. Los resultados obtenidos para diferentes conjuntos de rezagos considerados variando la cantidad de neuronas en la capa oculta son presentados en el Cuadro 3. El MLP-27 presenta el mejor ajuste a la muestra de calibración de parámetros; sin embargo, su SSE (0,033) es 235% superior al obtenido para el SVM-13 (0,014). Esta diferencia es igualmente notoria en términos del MAD.

Cuadro 3

**Sumatoria del error cuadrático medio (MSE) y desviación media absoluta (MAD):
perceptrón multicapa**

Modelo	Rezagos	H	Entrenamiento SSE (MAD)	Predicción SSE (MAD) 1 año	Predicción SSE (MAD) 2 años
MLP-1	1	1	0,191 (0,023)	0,010 (0,022)	0,020 (0,020)
MLP-2	1	2	0,190 (0,023)	0,010 (0,022)	0,019 (0,020)
MLP-3	1-2	1	0,190 (0,023)	0,010 (0,022)	0,019 (0,020)
MLP-4	1-2	2	0,189 (0,023)	0,010 (0,022)	0,020 (0,020)
MLP-5	1-3	1	0,175 (0,023)	0,012 (0,024)	0,023 (0,022)
MLP-6	1-3	2	0,175 (0,023)	0,012 (0,024)	0,023 (0,022)
MLP-7	1-4	1	0,170 (0,023)	0,011 (0,023)	0,020 (0,021)
MLP-8	1-4	2	0,170 (0,023)	0,011 (0,023)	0,020 (0,021)
MLP-9	1-5	1	0,169 (0,023)	0,010 (0,023)	0,021 (0,021)
MLP-10	1-5	2	0,169 (0,023)	0,011 (0,023)	0,020 (0,021)
MLP-11	1-6	1	0,131 (0,023)	0,014 (0,028)	0,026 (0,025)
MLP-12	1-6	2	0,130 (0,023)	0,014 (0,028)	0,026 (0,025)

Continúa

Modelo	Rezagos	H	Entrenamiento SSE (MAD)	Predicción SSE (MAD) 1 año	Predicción SSE (MAD) 2 años
MLP-13	1-7	1	0,130 (0,023)	0,014 (0,028)	0,026 (0,026)
MLP-14	1-7	2	0,130 (0,023)	0,015 (0,028)	0,027 (0,026)
MLP-15	1-8	1	0,118 (0,023)	0,013 (0,026)	0,023 (0,024)
MLP-16	1-8	2	0,118 (0,023)	0,013 (0,026)	0,023 (0,024)
MLP-17	1-9	1	0,118 (0,023)	0,013 (0,026)	0,023 (0,024)
MLP-18	1-9	2	0,118 (0,023)	0,013 (0,026)	0,023 (0,024)
MLP-19	1-10	1	0,107 (0,021)	0,014 (0,025)	0,040 (0,032)
MLP-20	1-10	2	0,098 (0,020)	0,013 (0,025)	0,033 (0,029)
MLP-21	1-10	3	0,065 (0,017)	0,012 (0,023)	0,023 (0,024)
MLP-22	1-11	1	0,106 (0,022)	0,014 (0,025)	0,039 (0,031)
MLP-23	1-11	2	0,105 (0,021)	0,014 (0,021)	0,038 (0,031)
MLP-24	1-11	3	0,101 (0,020)	0,013 (0,025)	0,032 (0,029)
MLP-25	1-12	1	0,106 (0,021)	0,015 (0,026)	0,040 (0,031)
MLP-26	1-12	2	0,074 (0,017)	0,021 (0,033)	0,045 (0,037)
MLP-27	1-12	3	0,033 (0,014)	0,016 (0,030)	0,037 (0,033)
MLP-28	1-13	1	0,074 (0,018)	0,014 (0,025)	0,029 (0,027)
MLP-29	1-13	2	0,074 (0,025)	0,015 (0,025)	0,032 (0,029)
MLP-30	1-13	3	0,033 (0,013)	0,018 (0,029)	0,061 (0,038)

Fuente: elaboración propia.

Los modelos MLP-1, MLP-2, MLP-3, MLP-4 y MLP-9 tienen el menor error para la predicción con horizonte de un año, pero su ajuste a la muestra de calibración es bastante inferior en calidad respecto a los demás modelos reportados en el Cuadro 3. Los modelos MLP-2 y MLP-3 también son los mejores en términos de ajuste a la muestra de predicción con un horizonte de dos años. Los modelos MLP-27 y MLP-30 tienen las menores SSE (de 0,049 y 0,051, respectivamente) al considerar el error sobre las muestras de

calibración (o entrenamiento en la jerga de inteligencia artificial) y de predicción un año adelante. Sin embargo, dichos modelos se ajustan pobremente al segundo año de predicción, de tal forma que sus SSE aumentan a 0,070 y 0,094; no obstante, el modelo MLP-27 es el que mejor se ajusta a los datos disponibles, en términos de los criterios de error considerados.

Un aspecto metodológico obligatorio en el desarrollo de modelos de predicción no

lineales lo constituye la comparación del modelo propuesto con la aproximación lineal tradicional. Ya que la serie presenta una clara componente de tendencia y un patrón cíclico de periodicidad anual, se procedió a transformarla aplicando los operadores de diferenciación simple y diferenciación estacional de período 12. Para pronosticar la nueva serie obtenida, se utilizaron modelos autorregresivos de orden P , con $P=1, \dots, 13$. Los estadísticos de ajuste a las muestras de calibración y predicción se presentan en el Cuadro 4.

La primera inspección de los resultados indica que aun cuando existe una diferencia

apreciable entre los ajustes de los distintos modelos a la muestra de calibración, esta se reduce bastante al considerar la predicción para uno y dos años de horizonte. Así es como el mayor SSE es el 26% y el 36% superior respecto al menor SSE para la correspondiente muestra de predicción. Tal como en los casos anteriores, se evaluó el ajuste de los modelos a las muestras de entrenamiento y de predicción, y se encontró que el modelo ARIMA-13 presenta el mejor ajuste a la totalidad de los datos.

En el Cuadro 5 se resumen los estadísticos calculados para los mejores modelos encontrados. En este caso, el modelo SVM-13

Cuadro 4

Sumatoria del error cuadrático medio (MSE) y desviación media absoluta (MAD): modelo ARIMA

Modelo	Rezagos	Entrenamiento SSE (MAD)	Predicción SSE (MAD) 1 año	Predicción SSE (MAD) 2 años
ARIMA-1	1	0,138 (0,023)	0,015 (0,031)	0,023 (0,026)
ARIMA-2	1-2	0,135 (0,023)	0,015 (0,031)	0,023 (0,026)
ARIMA-3	1-3	0,135 (0,023)	0,015 (0,031)	0,022 (0,026)
ARIMA-4	1-4	0,135 (0,023)	0,015 (0,031)	0,022 (0,026)
ARIMA-5	1-5	0,124 (0,023)	0,019 (0,033)	0,028 (0,027)
ARIMA-6	1-6	0,121 (0,023)	0,018 (0,033)	0,027 (0,028)
ARIMA-7	1-7	0,099 (0,021)	0,017 (0,033)	0,027 (0,029)
ARIMA-8	1-8	0,098 (0,021)	0,017 (0,034)	0,028 (0,029)
ARIMA-9	1-9	0,097 (0,021)	0,017 (0,034)	0,027 (0,029)
ARIMA-10	1-10	0,078 (0,019)	0,018 (0,035)	0,030 (0,031)
ARIMA-11	1-11	0,069 (0,018)	0,018 (0,034)	0,030 (0,030)
ARIMA-12	1-12	0,067 (0,018)	0,017 (0,033)	0,026 (0,028)
ARIMA-13	1-13	0,063 (0,018)	0,019 (0,034)	0,027 (0,027)

Fuente: elaboración propia.

presenta los mejores estadísticos de ajuste, con un SSE de 0,038, al considerar simultáneamente las muestras para entrenamiento y para predicción un año adelante; es seguido por el modelo MLP-27, que tiene un SSE de 0,049. Al tener en cuenta las muestras de calibración y de predicción 24 meses adelante, se obtiene que el modelo MLP-27 (con un SSE de 0,070) es el mejor predictor al considerar toda la información disponible.

En los gráficos 2 y 3 se presenta la predicción de los modelos SVM-13 y MLP-27 para un horizonte de 24 meses. En dichos gráficos, la serie real es representada como una secuencia de puntos; esta se distingue claramente en el extremo izquierdo y el extremo derecho del gráfico. Para ambos modelos se grafican su predicción puntual (valor esperado) y los intervalos de confianza de la predicción para un nivel de confianza del 95%. La predicción puntual es representada como una línea continua punteada, mientras que sus intervalos

asociados (que aparecen alrededor de ella) se notan únicamente como líneas continuas; estas tres líneas son claramente visibles en el extremo derecho de los gráficos durante la predicción para la muestra de validación.

Conclusiones

En este artículo se pronostican los precios promedios mensuales de los contratos despachados en el mercado eléctrico de Colombia, usando una máquina de vectores de soporte, un perceptrón multicapa y un modelo ARIMA. La predicción de esta serie es bastante compleja, debido a la presencia de cambios en la amplitud en su patrón cíclico de período anual, así como varios cambios en su tendencia de largo plazo durante el período analizado.

Las primeras 122 observaciones se usaron para estimar los parámetros de todos los modelos considerados, mientras que las res-

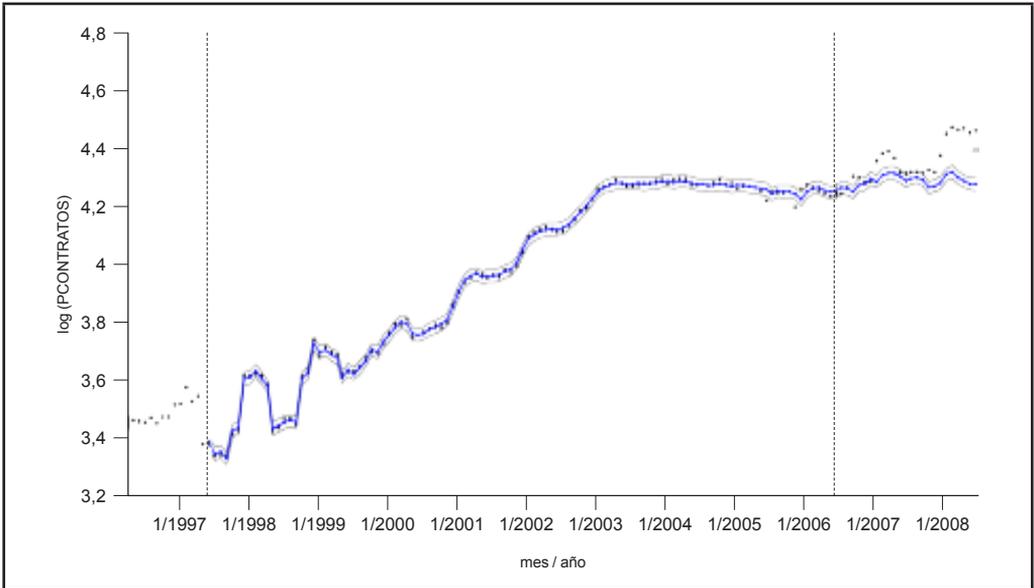
Cuadro 5

Sumatoria del error cuadrático medio (MSE) y desviación media absoluta (MAD): resumen

Modelo	Rezagos	Entrenamiento SSE (MAD)	Predicción SSE (MAD) 1 año	Predicción SSE (MAD) 2 años
SVM-13	1-13	0,014 (0,009)	0,024 (0,037)	0,210 (0,071)
MLP-27	1-12	0,033 (0,014)	0,016 (0,030)	0,037 (0,033)
MLP-30	1-13	0,033 (0,013)	0,018 (0,029)	0,061 (0,038)
SVM-17	1, 5, 8, 12, 13	0,047 (0,015)	0,014 (0,025)	0,105 (0,050)
SVM-11	1-11	0,048 (0,016)	0,017 (0,028)	0,133 (0,059)
SVM-15	1, 12, 13	0,076 (0,018)	0,012 (0,023)	0,052 (0,035)
SVM-2	1, 2	0,186 (0,022)	0,012 (0,024)	0,080 (0,043)
ARIMA-13	1-13	0,063 (0,018)	0,019 (0,034)	0,027 (0,027)

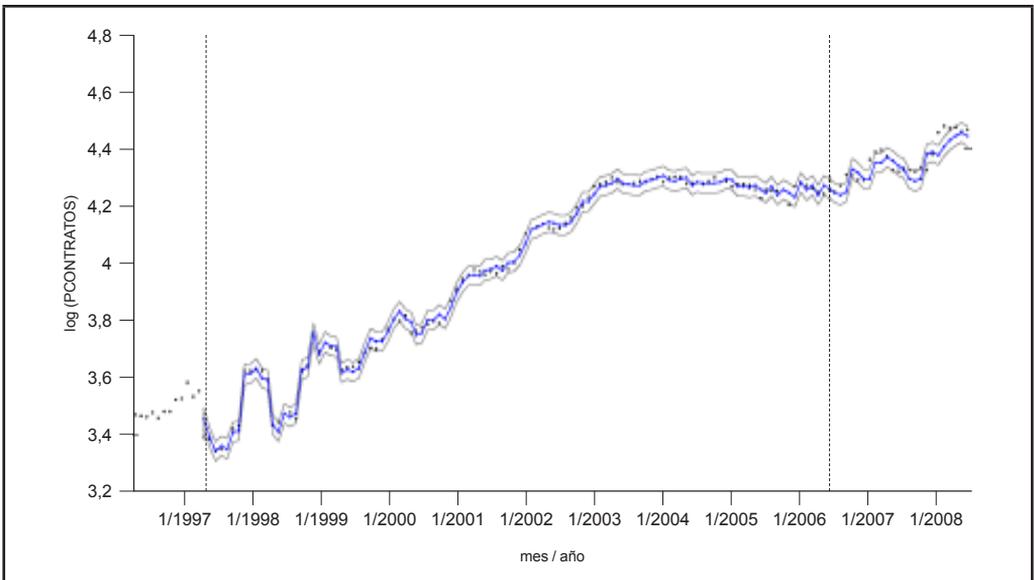
Fuente: elaboración propia.

Gráfico 2
Predicción obtenida usando del modelo SVM-13



Fuente: elaboración propia.

Gráfico 3
Predicción obtenida usando del modelo MLP-27



Fuente: elaboración propia.

tantes se utilizaron para evaluar su capacidad de predicción. Debido a la complejidad de la dinámica de la serie durante la muestra de predicción, se tomaron horizontes de 12 y 24 meses, que iniciaron siempre en la observación 123. Los resultados muestran que la SVM pronosticó de una forma más precisa el valor de la serie para el siguiente mes, tanto para la muestra de calibración como para el horizonte de 12 meses adelante; no obstante, un perceptrón multicapa presentó el mejor desempeño respecto a la totalidad de los datos. Este resultado se explica en parte, porque el MLP no capturó mejor la dinámica de la muestra de calibración en relación a la SVM, y ello le permitió (al MLP) responder mejor al cambio estructural que se presenta en la muestra de predicción.

En esta investigación no se han agotado todos los modelos disponibles, y ha sido realizada con la información disponible en el momento. Por ello es necesario revisarla a la luz de nueva información e incluir en la comparación otros modelos no lineales, como otras arquitecturas de redes neuronales, redes neurorodifusas y modelos estadísticos, como los mencionados en este artículo.

Lista de referencias

- Belousov, A. I.; Verzhakov, S. A. and von Frese, J. (2002). A flexible classification approach with optimal generalisation performance: support vector machines. *Chemometrics and Intelligent Laboratory Systems*, 64, 15-25.
- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2, 121-167.
- Chen, R. and Tsay, R. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, 88, 298-308.
- Franses, P. H. and van Djck, D. (2000). *Non-linear time series models in empirical finance*. Cambridge: Cambridge University Press.
- Gunn, S. R. (1997). *Support vector machines for classification and regression* (Technical report). Southampton, UK: Image Speech and Intelligent Systems Research Group, University of Southampton.
- Kaastra, I. and Boyd, M. (1996). Designing a neural network for forecasting financial and economic series. *Neurocomputing*, 10, 215-236.
- Masters, T. (1993). *Practical neural network recipes in C++*. New York: Academic Press.
- (1995). *Neural, novel and hybrid algorithms for time series prediction*. New York: John Wiley and Sons.
- Mukherjee, S.; Osuna, E. and Girosi, F. (1997). Non-linear prediction of chaotic time series using support vector machines. Documento presentado en IEEE NNSP'97, Amelia Island, FL, Estados Unidos.
- Muller, K. R.; Smola, A.; Ratsch, G.; Scholkopf, B.; Kohlmorgen, J. and Vapnik, V. N. (1997). Predicting time series with support vector machines. En: W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud (Eds.), *Proceedings of ICANN '97*, Springer LNCS 1327, Berlin, 999-1004.
- Osowski, S. and Garanty, K. (2007). Forecasting of the daily meteorological pollution using

- wavelets and support vector machine. *Engineering Applications of Artificial Intelligence*, 20, 745-755.
- Platt, L. (1998). Fast training of SVM using sequential optimization. En: B. Scholkopf, B. Burges y A. Smola (Eds.), *Advances in Kernel methods—support vector learning* (pp. 185-208). Cambridge: MIT Press.
- Scholkopf, B. and Smola, A. (2002). *Learning with Kernels*. Cambridge, MA: MIT Press.
- Tay, F. E. H. and Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29, 309-317.
- Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89, 208-218.
- ; Lin, C. and Granger, C. (1993). Power of the neural network linearity test. *Journal of Time Series Analysis*, 14, 209-220.
- Thissen, U.; van Brakel, R.; de Weijer, A. P.; Melsen, W. J. and Buydens, L. M. C. (2003). Using vector support machines for time series prediction. *Chemometrics and Intelligent Laboratory Systems*, 69, 35-49.
- Tong, G. (1978). On a threshold model. En: C. Chen (Ed.), *Pattern recognition and signal processing*. Amsterdam: Sijhoff & Noordhoff.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer.
- (1998). *Statistical learning theory*. New York: Wiley.
- Vapnik, V. N.; Golowich, S. E. and Smola, A. J. (1996). Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems*, 9, 281-287.
- Velásquez, J. D.; Dyner, I. y Souza, R. C. (2007). ¿Por qué es tan difícil obtener buenos pronósticos de los precios de la electricidad en mercados competitivos? *Cuadernos de Administración*, 20 (34), 259-282.
- Zhang, G.; Patuwo, B. and Hu, M. (1998). Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, 14, 35-62.

