IJEEP

INTERNATIONAL JOURNAL OF ENERGY ECONOMICS AND POLICY

EJ EconJournals

ISSN: 2146-4553

International Journal of Energy Economics and Policy

ISSN: 2146-4553

available at http: www.econjournals.com

International Journal of Energy Economics and Policy, 2021, 11(2), 67-74.

EJ EconJournals

Forecasting the Colombian Electricity Spot Price under a Functional Approach

Santiago Gallón, Jorge Barrientos*

[1]Departamento de Matemáticas y Estadística, Facultad de Ciencias Económicas, Universidad de Antioquia, Medellín, Colombia,
[2]Departamento de Economía, Facultad de Ciencias Económicas, Universidad de Antioquia, Medellín, Colombia.
*Email: jorge.barrientos@udea.edu.co

Received: 14 September 2020 **Accepted:** 19 December 2020 **DOI:** https://doi.org/10.32479/ijeep.10607

ABSTRACT

Forecasting the hourly electricity spot price plays a crucial role for agents involved in energy day-ahead markets. However, traditional time series processes used for this issue model each hour separately not taking into account the intraday energy market microstructure information. In this paper, we appeal to a Functional Data Analysis (FDA) viewpoint that allows modeling and forecasting the intraday electricity spot price of the Colombian Electricity Market. Specifically, we use the Hyndman-Ullah-Shang method, which relies on a functional principal component decomposition of the nonparametric smoothed price curves, where the short-term forecasts are obtained by using the empirical functional principal components and the univariate time series forecasts of the corresponding estimated scores. Results show that one of the main advantages of this approach is that it allows to capture the underlying intraday common structural patterns shared by the daily spot price curves, and also behaves well for one-month-ahead price predictions compared with standard benchmarks.

Keywords: Day-ahead Electricity Price Forecasting, Functional Data Analysis, Functional Principal Components, Functional Time Series Forecasting
JEL Classifications: C32, C53, C55, Q41, Q47

1. INTRODUCTION

It is well-known that due to the worldwide deregulation process and structural reforms carried out during the 1990s, which brought significant changes in electricity markets, such as considering the electrical energy as a commodity trading upon competitive market rules (Huisman et al., 2007; Weron, 2006), there has been a growing interest by understanding the underlying dynamic structure of the data generating process of the electricity spot price for competitive pool-based electricity markets, and accurately forecasting it at short-, medium- and long-terms (Conejo et al., 2005; Weron, 2006; 2014). In Colombia, these reforms were established with the Domiciliary Public Services and Electricity Laws 142 and 143 (Congress of the Republic of Colombia, 1994), creating conditions to ensure an efficient energy supply under social, economic, environmental and financial feasibility criteria, and to avoid the abuse of dominant positions. As a result, the Colombian market became one of the most open markets among the developing countries.

Hence, electricity spot price forecast has become crucial for agents involved in energy electricity markets, specially, distribution companies as well as generators and traders, owing to the price may contain information to anticipate decisions related with the installed capacity needed to cope with the energy demand, minimize investment risks associated to the price volatility, design contracts at different maturities and it helps generation companies in planning their bidding strategies in order to maximizing profits in the short term (Conejo et al., 2005; Shahidehpour et al., 2002; Weron, 2014; Panchakshara et al., 2014). It is particularly important for the Colombian case because approximately 80% of energy is traded through forward contracts.

This task has imposed big statistical challenges mainly due to the (1) non-storable nature of electricity, (2) high intraday energy demand variability, (3) variable costs are low compared to the large fixed costs, (4) high dependence on the hydrological component, (5) high price uncertainty, and (6) lack of regulation that forces electricity-generating firms to report their variable costs to avoid market power and information asymmetries. Thus, the spot price has empirical regularities not easily captured by statistical time series models designed to storable commodities, such as (1) superposed seasonal patterns, (2) mean-reversion, (3) presence of sudden and unexpected spikes, and (4) volatility (Aggarwal et al., 2009; Conejo et al., 2005; Crespo et al., 2004; Huisman et al., 2007; Liebl, 2013; Weron, 2006; 2014).

In the statistical literature, there is a wide range of methods to model and forecast the electricity spot price (see Aggarwal et al., 2009; Weron, 2006, 2014; Weron and Misiorek, 2008) for reviews about the subject). Some of these are, for instance, exponential smoothing filters, Autoregressive Integrated Moving Average with exogenous terms (ARIMAX) type processes, seasonal and long-memory models, Autoregressive Conditional Heteroscedasticity (ARCH) class of processes, state space and unobserved component models, mean-reversion jump-diffusion models, Markov regime switching processes, semiparametric and nonparametric methods, among others. Also, machine learning algorithms such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Random Forests (RFs) have been applied.

In Colombia, there are few papers dealing with electricity spot price forecasting. Lira et al. (2009) forecasts the daily price considering ARMAX and Periodic-ARMAX (PARMAX) processes including oil/gas prices, water reservoir levels, river contribution, and load demand as exogenous variables. Also ANNs and fuzzy algorithms are used. Likewise, Barrientos et al. (2012; 2018) apply ARMAX, Vector Error Correction (VEC) and Non-linear Autoregressive Neural Networks (NARX) models based on reservoir levels, load demand, and energy supply to predict the monthly electricity price. On the other side, Bello and Beltrán (2010) models and forecasts the daily energy price with ARCH family models.

However, in electricity spot markets the hourly price is determined in a day-ahead blind auction through an Independent System Operator (ISO) trading 24/7 without exceptions in which agents submit their hourly bids of the electricity price and physical power deliveries for the next day. Thereby, the 24 intra-day prices for each day are settled jointly the previous day. It implies that hourly electricity spot prices cannot be viewed as a time series process (Huisman et al., 2007), as is assumed in all of the above classical time series models, and ignoring the intraday market microstructure dynamics present in the day-ahead energy markets.

To tackle this day-ahead forecasting issue, we adopt a Functional Data Analysis (FDA) approach (Horváth and Kokoszka, 2012; Kokoszka and Horváth, 2017; Ramsay and Dalzell, 1991; Ramsay and Silverman, 2005; Wang et al., 2016, and the references given there) by considering the hourly spot price time series per day as a collection of curves. It is, the sample of random variables is assumed as a set of functions belonging to some infinite-dimensional space F instead of a finite-dimensional one as is usual in most statistical applications. Hence, the variables are viewed as paths of a smooth stochastic processes $Y=\{Y(x):x \in \chi \subset R\} \in F$ defined on some index set χ, where the observed dataset is obtained from discretizations at points x_1,..., x_J Usually, this approach is appropriate in cases including irregularly spaced measurements, high-frequency data, sparsely observed curves, analysis upon derivatives of functions, among others.

Although the field of FDA is relatively new, the number of disciplines where it has been applied is huge because of technological advances that allow to collect and store high-dimensional data observed continuously on an interval or intermittently at several points (Wang et al., 2016). Applications are in bioinformatics, medicine, finance, ecology, meteorology, demography, etc. Despite its success, there are few studies on modeling and forecasting the electricity price, and also the demand, with this setting (Aneiros et al., 2016; Liebl, 2013; Portela et al., 2018; Shang, 2013; Vilar et al., 2012), and there are no empirical studies for the Colombian case.

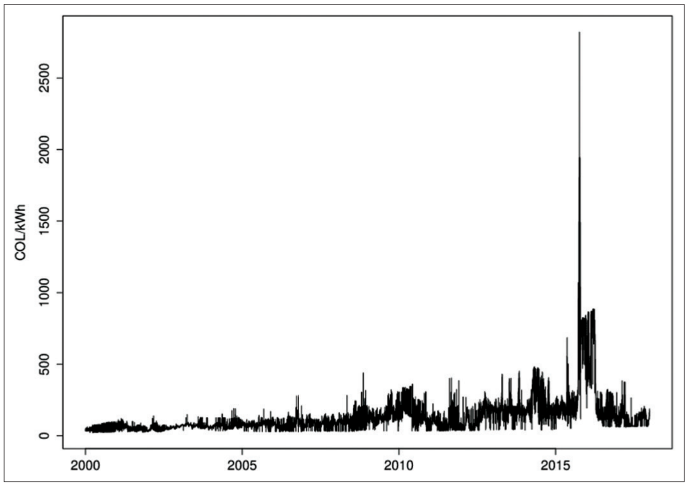
Here, we apply the model by Hyndman and Ullah (2007); Hyndman and Shang (2009), which is based on the functional principal component decomposition of the smoothed curves, where the forecasts are obtained by using the sample functional principal components, and the h step-ahead forecasts of the respective scores by fitting a univariate time series model for each of them. It has been successfully applied to forecast variables ranging from mortality/fertility rates to intraday index returns. Results show that it is possible to capture intraday common features present in the price curves, and behaves well for short- and also medium-term forecasts.

The paper is organized as follows. Section 2 presents the electricity spot price data, and explains how the functional data is obtained. Section 3 describes the used functional time series forecast method by Hyndman and Ullah (2007); Hyndman and Shang (2009). Results for hour-specific day-ahead and one-month-ahead price predictions are shown in Section 4. Last, Section 5 concludes.

2. DATA

The data shown in the Figure 1 corresponds to the hourly electricity spot price (Colombian Pesos per kilowatthour, COP/kWh) of the Colombian Electricity Market from January 1, 2000 (Saturday) to December 31, 2017 (Sunday), $\{Y_\tau, \tau \in [1,T]\}$, T = 157800, available at XM (http://www.xm.com.co/), an affiliate of *Interconexión Eléctrica S.A.* –ISA- (http://www.isa.co/). In normal operating conditions, the price is the higher one offered by generators that have been programmed to cover the ideal energy dispatch, and represents an unique price given by the marginal cost for the Energy Interconnected System at each time point (Huisman et al., 2007; Shahidehpour et al., 2002; Weron, 2014).

Figure 1: Colombian hourly electricity spot price, January 1, 2000-December 31, 2017



2.1. Functional Data

Owing to the 24 intra-day structure of electricity spot prices vary substantially across days (Huisman et al., 2007; Liebl, 2013; Weron, 2006), the time series can be splitted into daily datasets, $\{Y_{i,\tau}, \tau = 1,\ldots,T_i\}$, i = Mon.,…, Sun., where T denotes the number of time data points of the *i* th day. To follow the FDA approach, the samples of hourly time series curves per day were obtained by converting the T_i time points into $n_i = T_i/J$ daily functions defined on $x \in [0, J]$ with J = 24 hours (Shang, 2013),

$Y_{i,t}(x) = \{Y_{i,\tau}, \tau \in [J(t-1), Jt]\}$ $t=1,\ldots,n_{\_i}$, for i = Mon.,…, Sun.,

where the observations are assumed to be discretizations generated generated by evaluating, with error $\varepsilon_{i,t,j}$ a set of unknown smooth functions $f_{i,t}(x)$ at points x_j satisfying the model,

$Y_{i,t}(x_j)$ = f_{i,t}(x_j)+\varpesilon_{i,t,j}), j=1,…,J, $\varepsilon_{i,t,j}$ is i.i.d (0,1) (1)

The sample paths of $f_{i,t}(x)$ for each *i* and *t* were estimated from the observed pairs $\left\{\left(x_j, y_{i,t}\left(x_j\right)\right)\right\}_{j=1}^{J}$ by applying some nonparametric curve estimation method. We applied smoothing B-splines using the fda package (Ramsay et al., 2017) in the open-source R system for statistical computing (R Development Core Team, 2018). Smoothing spline finds $\hat{f}_{i,t}$ such that minimizes the penalized sum of squared errors[1],

$$\sum_{j=1}^{J}\left(y_{i,t}\left(x_j\right) - f_{i,t}\left(x_j\right)\right)^2 + \lambda_i \int_0^J \left(f_{i,t}''(\tau)\right)^2 d\tau$$

with $f_{i,t}(x) = \sum_{l=1}^{L} \theta_{i,t,l}\varphi_{i,t,l}(x)$ following a truncated basis expansion, and the regularization parameter λ_i controls the amount of smoothing, chosen by generalized cross-validation (Ramsay and Silverman, 2005).

Due to the high dimensionality of functional data, it is probable that some curves attain extreme values at a single or several points, or have remarkable different shapes from the rest or curves, or both. It can have serious effects on modeling and forecasting tasks leading to erroneous conclusions. To detect functional outliers, the bivariate and functional Highest Density Region (HDR) boxplots proposed by Hyndman and Shang (2010) were applied on $\hat{f}_{i,t}$. The bivariate HDR boxplot is based on the HDR defined as $R_\alpha = \left\{z : \hat{f}(z) \geq f_\alpha\right\}$, where $\hat{f}(z)$ is the bivariate kernel density estimate of the first two principal component scores (Appendix I) of the smoothed functions, and f_α is such that $\Pr(\boldsymbol{Z} \in R_\alpha) \geq 1 - \alpha$. Points within R_α have higher density than those lying outside. The bivariate HDR boxplot displays the point corresponding to the mode curve (i.e. the highest density point), the 50% inner (the "bag") and 99% outer (the "fence") HDRs, and points outside the outer HDR which are flagged as potential outliers (Hyndman and Shang, 2010, for details). An example of this plot for the sample of curves on Tuesdays using the R package rainbow (Shang and Hyndman, 2016) is shown in the Figure 2 (on the left)[2]. The corresponding functional HDR boxplot (on the right) is the mapping of the bivariate HDR boxplot to the functional curves, where the dark (light) gray region shows the 50% (99%) inner (outer) HDR.

The detected outliers were mainly those days in which the price was significantly high, coinciding with the "El Niño" phenomenon characterized by warm and wet weather months of April-July during 2014-2016. Other atypical curves of 2009-2011 corresponded to days of the last quarter of 2017, marked by a high energy consumption season reflected on price increases. For the rest of days, the functional outliers, in general, corresponded to similar dates. Figure 3 shows the resulting sample smoothed electricity price curves $\hat{f}_{i,t}$ without the identified outliers for Tuesdays. We can observe that curves capture the common structural patterns which characterize the hourly electricity price variability. It is, the shape of curves shows that from approximately 5:00 a.m. the spot price tends to increase progressively until 12:00 p.m., reaches a second peak around 8:00 p.m., and after begins to decrease progressively.

3. FUNCTIONAL TIME SERIES

3.1. Functional Principal Components

To forecast the electricity price curves for each day, we adopt the functional principal component approach by Hyndman and Ullah (2007), Hyndman and Shang (2009), and Shang (2010) based on the Karhunen-Loève decomposition (Appendix I), i.e. from a basis function expansion for each *i*th day,

$$f_{i,t}(x) = \mu_i(x) + \sum_{k=1}^{K_i} \beta_{i,t,k}\phi_{i,k}(x) + \xi_{i,t}(x), \quad K_i < n_i, \qquad (2)$$

where μi (x) is the i[th] mean function, $\left\{\phi_{i,k}(x)\right\}_{k=1}^{K_i}$ the set of orthonormal basis (principal components) functions with corresponding dynamic coefficients (scores) $\left\{\beta_{i,t,k}\right\}_{k=1}^{K_i}$, K_i the number of basis, and $\left\{\xi_{i,t}(x)\right\}_{t=1}^{n_i}$ are centered i.i.d. random

1 The solution $\hat{f}_{i,t}$ is a natural cubic spline with knots $\{x_1,\ldots,x_J\}$.

2 As an illustration of the obtained results and to save space, all figures in the paper are presented for Tuesdays. Figures for other days of the week are available upon request.

Figure 2: Bivariate (left) and functional (right) HDR boxplots of spot price on Tuesdays. Dark and light gray regions show the bag and fence HDRs, resp. The black line is the modal curve. Numbers (on the left) and corresponding colored curves (on the right) outside the fence identify atypical days whose list are available upon request

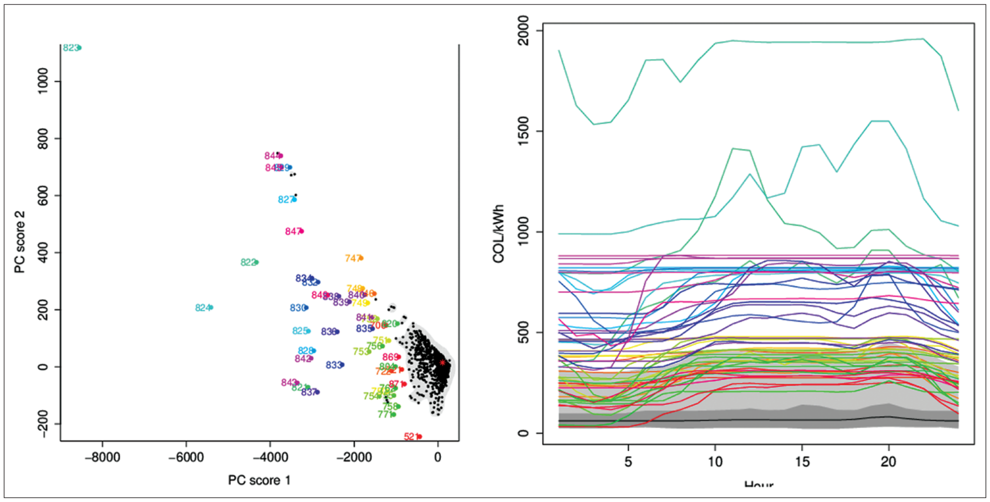
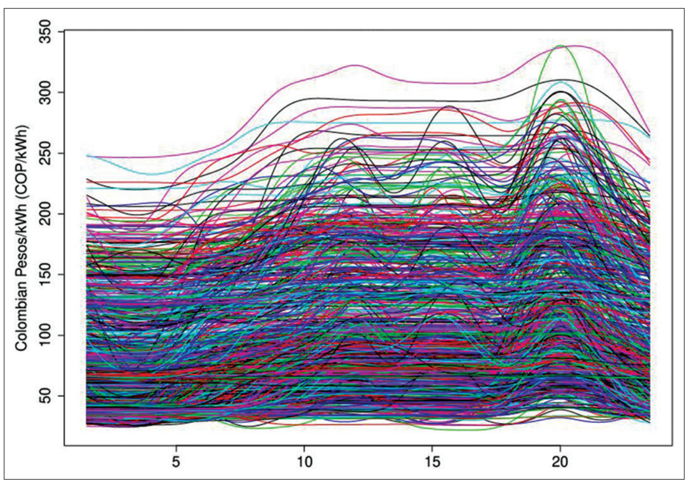


Figure 3:. Electricity price curves $\hat{f}_{i,t}$ on Tuesdays after removing functional outliers.



functions. To estimate $\mu_i(x)$, we used the weighted average proposal by Hyndman and Shang (2009), $\hat{\mu}_i(x) = \sum_{t=1}^{n_i} \omega_{i,t} \hat{f}_{i,t}(x)$, where $\omega_{i,t}$ are weights assigning more weight to recent observations.

There are several computational methods to obtain the sample basis functions and scores (Hyndman and Ullah, 2007; Ramsay and Dalzell, 1991; Ramsay and Silverman, 2005; Shang, 2014). We applied the discretization-based approach which is based on the singular vale decomposition $F_i = \Phi_i \Lambda_i V_i$ of the $q \times n_i$ centered matrix F_i obtained from a dense equally spaced discretization $\{x_1, \dots, x_q\} \in [0,J]$ of the sample centered functions $\hat{f}_{i,t}(x_j) - \hat{\mu}_i(x_j)$, where the $(j,k)^{th}$ basis $\hat{\phi}_{i,k}(x_j)$ is the $(j,k)^{th}$ coordinate of Φ_i, and the respective score $\hat{\beta}_{i,t,k}$ is the $(j,k)^{th}$ coordinate of $F_i^T \Phi_i$ (Hyndman and Shang, 2009; Ramsay and Dalzell, 1991). Figure 4 illustrates an example of the functional principal component expansion on the price curves on Tuesdays using the R package ftsa (Hyndman and Shang, 2017) with K=1 obtained by following the method of Shang (2013) described below.

3.2. Functional Forecasts

With estimates $\left\{\left\{\hat{\mu}_i(x)\right\}, \hat{\phi}_{i,k}(x), \hat{\beta}_{i,t,k}\right\}$, and from (1) – (2), the model reduces to

$$y_{i,t}(x_j) = \hat{\mu}_i(x) + \sum_{k=1}^{K_i} \hat{\beta}_{i,t,k} \hat{\phi}_{i,k}(x_j) + \hat{\xi}_{i,t}(x_j) + \hat{\varepsilon}_{i,t,j}$$

Due to orthonormality of basis functions, the scores $\left\{\hat{\beta}_{i,t,k}\right\}_{k=1}^{K_i}$ can be independently forecasted with, for instance, an ARIMA process (Hyndman and Ullah, 2007; Hyndman and Shang, 2009; Shang, 2013). Thus, conditional to $\left\{y_{i,t}(x_j) : t = 1, \dots, n_i,\ j = 1, \dots, J\right\}$, and $\left\{\hat{\phi}_{i,k}(x) : k = 1, \dots, K_i\right\}$, the forecasts are

$$\hat{Y}_{i,n_{i+h|n_i}}(x) = \hat{\mu}_i(x) + \sum_{k=1}^{K_i} \hat{\beta}_{i,n_i+h|n_i,k} \hat{\phi}_{i,k}(x), \tag{3}$$

where $\hat{\beta}_{i,n_i+h|n_i,k}$ is the -step-ahead forecast of $\beta_{i,n\_i+h,k}$.

The optimal number of principal components K_i was chosen with the validation method of Shang (2013) in which the sample of curves for each day is splitted into a training set with $n_i^* = n_i - l$ functions, and a validation set with l=52 curves corresponding to weeks of the year. Then, an accuracy measure is calculated with the forecasts for functions in the validation set based on the functional model fitted with the training set. For a potential number of basis K_i=1,…,10, Table 1 reports the Mean Absolute Percentage Error (MAPE), mean $\left(|\, p_{h,j} \,|\right)$, where:

$$p_{h,j} = \frac{y_{i,n_i^*+h}(x_j) - \hat{Y}_{i,n_i^*+h|n_i^*}(x_j)}{y_{i,n_i^*+h}(x_j)} \cdot 100\%.\ h = 1, \dots, l,\ j = 1, \dots, J$$

Figure 4: Mean function, first functional principal component and score on Tuesdays

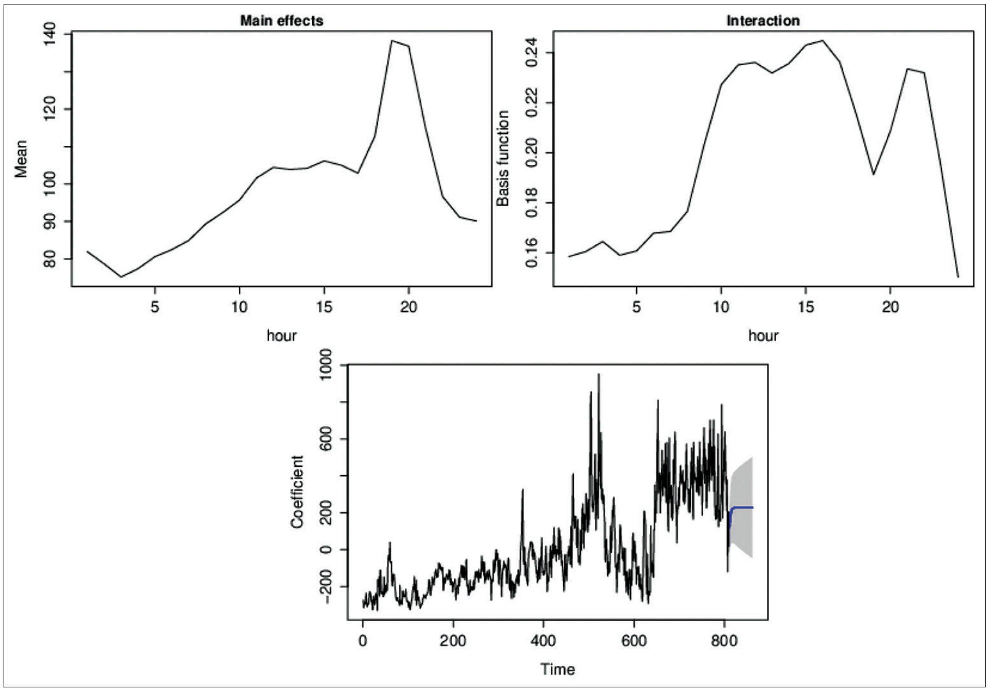


Table 1: Mean absolute percentage error

K	Mondays	Tuesdays	Wednesdays	Thursdays	Fridays	Saturdays	Sundays
1	**46.574**	**51.017**	66.707	**51.269**	**50.119**	**55.113**	43.750
2	46.637	51.722	**66.547**	51.379	50.425	55.199	43.739
3	46.634	51.697	66.550	51.373	50.384	55.209	43.738
4	46.662	51.697	66.562	51.382	50.373	55.193	43.727
5	46.666	51.703	66.566	51.383	50.386	55.193	43.729
6	46.662	51.693	66.570	51.380	50.386	55.198	43.729
7	46.660	51.702	66.563	51.377	50.386	55.198	**43.724**
8	46.661	51.693	66.562	51.377	50.384	55.196	43.724
9	46.662	51.693	66.563	51.374	50.384	55.198	43.727
10	46.658	51.694	66.569	51.376	50.382	55.198	43.727
Prop.	0.9779	0.974	0.9914	0.9821	0.9838	0.9858	0.9993

Prop: proportion of variation explained by the optimal first principal components (in boldface)

The k_i ranges between 1 and 2, except to Sundays. Last row shows the proportion of variation explained by the optimal k_i first principal components varying between 97% and 99%[3]. Results were obtained with the R package ftsa (Hyndman and Shang, 2017).

4. FORECAST RESULTS

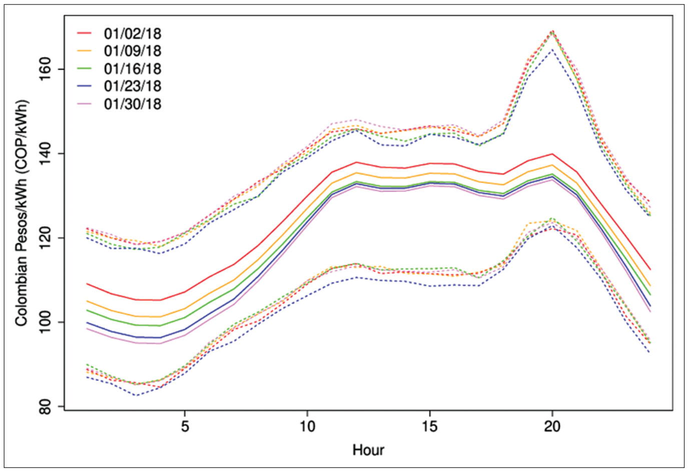
With the historical data of electricity spot price from January 1, 2000 to December 31, 2017, the optimal number of basis functions K_i, the corresponding empirical principal components $\left\{\hat{\phi}_{i,k}\left(x\right)\right\}_{k=1}^{K_i}$, and h-step-ahead forecasts $\hat{\beta}_{i,n_i+h|n_i,k}$, we obtain 24 hourly-specific h day-ahead forecasts according to (3). Although the main interest of the paper focuses on one day-ahead (h=1), forecasts we also carry out one-month-ahead price predictions for January 2018, where h=5 for Mondays, Tuesdays and Wednesdays, and h=4 for Thursdays, Fridays, Saturdays and Sundays for this particular month. The reason for medium-term predictions obeys the constant need by agents (mainly price-taker producers and retailers) in the Colombian Electricity Market to its balance sheet estimates, to engage bilateral agreements, and risk management.

As illustration, the Figure 5 shows the point forecasts for Tuesdays with its respective 80% interval forecasts using the nonparametric bootstrap method proposed by Hyndman and Shang (2009) and Shang (2013), with B=1000 bootstrap samples. We can see that forecasts exhibit the common structural features of the hourly electricity price variability described in Subsection 2.1. Besides, as it was expected, the spot price curve forecasts reveal a time-trend pattern in the sense that these follow a time-step-ahead horizon ordering. It is, an ordering from a greater predicted price for the most recent step-ahead horizon (shown in red) to a lower forecasted price for the most distant step-ahead horizon (shown in violet).

Forecasts were compared with those from functional and non-functional bench-marks. We used a Functional AutoRegressive (FAR) process of order one (Besse and Cardot, 1996),

3 Similar results were got with the Root Mean Square Percentage Error (RMSPE), [mean $\left(p_{h,j}^2\right)$]$^{1/2}$.

Figure 5: Forecasts of hourly electricity spot price on Tuesdays for January 2018 with 95% nonparametric prediction intervals



$y_{i,t}(x)-\mu(x)=\rho\left(y_{i,t-1}(x)-¼(x)\right)+\mu_{i,t}$, with ρ a bounded linear autoregressive operator on a Hilbert space H, $\mu(x) \in H$, and $\varepsilon_{i,t}$ a sequence of i.i.d. zero-mean errors in H. It was fitted with the package far (Damon and Guillas, 2015). As non-functional alternatives, we used the seasonal multiplicative ARIMA (SARIMA) and Neural Network Autoregressive (NNAR) models, which are two well-known processes used in the energy price forecasting literature (Weron, 2014). The SARIMA $(p,d,q)\times(P,D,Q)_s$ process is given by $\Phi_P\left(B^s\right)\phi_p(B)(1-B)^d\left(1-B^s\right)^D Y_t=\theta_q(B)\Theta_Q\left(B^s\right)\varepsilon_t$, where ϕ_p and Φ_P are ordinary and seasonal autoregressive p^{th} and P^{th} degree polynomials, θ_q and Θ_Q ordinary and seasonal moving-average q^{th} and Q^{th} degree polynomials, resp., B the backward shift operator $\left(B^k z := z_{t-k}\right)$, and s the seasonal period. The identified model was a SARIMA $(1,1,0)\times(1,1,0)_{168}$, $s=168=24 \times 7$. The fitted NNAR was single-hidden-layer NNAR$(p,P,k)_s$ with $p=24$ ordinary and $P=1$ seasonal autoregressive inputs, and $k=13$ nodes, given by $y_t=\pm_0+\sum_{j=1}^{k}\pm_j f_j\left(x_t^T w_j\right)+\mu_t$, with $x_t=\left(1,y_{t-1},\ldots,y_{t-p},y_{t-s}\right)$, and $f_j(\cdot)$ the sigmoid activation function. The reason for a seasonal components obeyed to the hourly periodic pattern found in the autocorrelation function.

We also considered two models of the exponential smoothing family: the Double-Seasonal Holt-Winters (DSHW) model (Taylor, 2003), which is an extension of the Holt-Winters method to handle high-frequency multiple seasonal patterns, and the exponential smoothing state space model with Trigonometric Box-Cox transformation, ARMA errors, Trend and Seasonal (TBATS) components (De Livera et al., 2011), which in turn is a generalization of the DSHW model. All non-functional methods were fitted with the package forecast (Hyndman et al., 2018).

According to MAPE results in the Table 2,[4] the Functional Time Series (FTS) model performs better than benchmarks for both short-term (one-day-ahead) and medium-term (one-month-ahead) forecasts. For the FTS model, the errors oscillated between 4.1% and 8.9% (6.7% in average) for short-term forecasts, and between

Table 2: MAPE for 1-day and 1-month ahead forecasts

Day	One-day ahead forecasts					
	FTS	FAR	SARIMA	NNAR	DSHW	TBATS
Mondays	**6.02**	7.71	12.86	14.64	13.31	9.91
Tuesdays	8.90	**7.67**	17.33	9.19	14.00	10.00
Wednesdays	**8.37**	8.73	18.30	17.9	11.85	9.56
Thursdays	6.90	**4.23**	10.27	5.47	8.57	5.16
Fridays	3.50	6.95	11.84	12.05	10.99	**3.24**
Saturdays	8.87	14.08	19.23	19.94	19.61	17.8
Sundays	**4.10**	10.73	4.19	11.04	7.83	5.39
Mean	**6.72**	8.53	12.86	14.64	13.31	9.91
Day	**One-month ahead forecasts**					
	FTS	FAR	SARIMA	NNAR	DSHW	TBATS
Mondays	10.28	10.27	15.42	19.02	18.94	**9.84**
Tuesdays	8.47	**7.24**	12.4	17.43	15.27	9.11
Wednesdays	**6.54**	8.33	10.89	18.77	8.23	8.09
Thursdays	**6.83**	10.53	9.65	22.75	25.37	13.05
Fridays	**7.04**	14.08	14.44	20.14	19.78	9.79
Saturdays	**10.13**	19.72	29.36	24.08	37.57	32.19
Sundays	**12.82**	23.60	18.62	18.53	25.06	15.75
Mean	**8.87**	13.40	15.42	19.02	18.94	9.84

Minimum MAPE values in boldface

Table 3: DM test statistics for the forecast accuracy of the FTS model versus FAR, SARIMA, NNAR, DSHW, and TBATS benchmarks

Day	Absolute-error loss				
	FAR	SARIMA	NNAR	DSHW	TBATS
Mondays	−1.85	−4.03	−1.77	−6.43	−3.50
	(0.03)	(0.00)	(0.04)	(0.00)	(0.00)
Tuesdays	0.71	−1.75	−0.92	−6.06	−0.64
	(0.76)	(0.04)	(0.18)	(0.00)	(0.26)
Wednesdays	−1.63	−3.61	−3.67	−1.64	−2.00
	(0.05)	(0.00)	(0.00)	(0.05)	(0.02)
Thursdays	−1.00	−2.07	−1.88	−2.41	−1.44
	(0.16)	(0.02)	(0.03)	(0.01)	(0.08)
Fridays	−2.06	−4.07	−3.42	−1.75	−1.51
	(0.01)	(0.00)	(0.00)	(0.04)	(0.07)
Saturdays	−5.24	−3.64	−8.78	−3.73	−2.94
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Sundays	−4.50	−1.80	−3.44	−2.58	−3.05
	(0.00)	(0.04)	(0.00)	(0.01)	(0.00)
	Squared-error loss				
	FAR	SARIMA	NNAR	DSHW	TBATS
Mondays	−1.63	−4.97	−3.37	−3.27	−3.30
	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)
Tuesdays	0.76	−2.23	−1.33	−6.23	−1.00
	(0.77)	(0.01)	(0.09)	(0.00)	(0.16)
Wednesdays	−1.66	−2.45	−2.93	−1.68	−1.83
	(0.05)	(0.00)	(0.00)	(0.05)	(0.03)
Thursdays	−1.15	−1.74	−1.57	−2.04	−1.48
	(0.13)	(0.04)	(0.06)	(0.02)	(0.07)
Fridays	−2.15	−2.78	−2.23	−1.57	−1.48
	(0.02)	(0.00)	(0.01)	(0.04)	(0.07)
Saturdays	−3.39	−2.29	−4.21	−2.42	−2.18
	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)
Sundays	−2.21	−1.71	−2.35	−1.87	−2.01
	(0.01)	(0.04)	(0.01)	(0.03)	(0.02)

p-values in parenthesis

4 Similar results were obtained by using other forecast performance metrics.

6.5% and 12.82% (8.9% in average) for medium-term forecasts. Among benchmarks, the FAR model was the best competitor of the FTS process followed by the TBATS model. In general, the worst forecaster was the NNAR model, followed by DSHW and SARIMA processes. Also for Saturdays and Sundays the one-month-ahead forecast errors were bigger than in the weekdays; this result is also found in Aneiros et al., (2016) for the electricity market of mainland Spain.

Besides, to evaluate the one-day ahead forecast accuracy of the FTS model with respect to the benchmarks over one-month out-of-sample window, the Diebold-Mariano (DM) test (Diebold and Mariano, 2002) was used. The null hypothesis is that the FTS model compared with the benchmarks have the same forecast accuracy, $H_0 : \mathrm{E}(d_t) = 0$, where $d_t = g(e_{FTS.t}) - g(e_{l,t})$, l = FAR, SARIMA, NNAR, DSHW, TBATS, is the loss differential between both forecasts with g(e) some loss function for the forecast error e. The alternative hypothesis is that the forecasts of the FTS method are more accurate than the benchmarks, $H_1 : \mathrm{E}(d_t) < 0$. The test statistic is $DM = \bar{d} / \hat{\sigma} \xrightarrow{D} \mathrm{N}(0,1)$, where $\bar{d}$ is the mean of d_t and $\hat{\sigma}$ the consistent sample standard deviation of $\bar{d}$. Table 3 reports the test statistic values using the absolute- and squared-error loss functions, and the respective *p*-values. Results show the FTS model has a statistical significant predictive behavior for forecasting the electricity spot price in most of cases, rejecting the null hypothesis at 5% level of significance, except to the FAR process for Tuesdays and Thursdays, and the TBATS model for Tuesdays, Thursdays and Fridays.

5. CONCLUDING REMARKS

In this paper we applied the functional time series model proposed by Hyndman and Ullah (2007); Hyndman and Shang (2009), which is based on a weighted functional principal component analysis, for forecasting the hour-specific *h*-ahead electricity spot price from the Colombian Electricity Market. The functional approach of the model allows to capture the intraday market microstructure dynamics in the day-ahead energy market in which the 24 intra-day prices for each day are established simultaneously the previous day, issue that is ignored in the classical time series models.

Specifically, we obtained one-day-ahead, and also one-month-ahead spot price curve forecasts. Results showed that out-of-sample predictions are able to pick up the intraday common structural characteristics in the spot price, and also have superior forecast performance in comparison with the FAR, SARIMA, NNAR, DSHW, and TBATS benchmarks, with Mean Absolute Percentage Errors of 6.7%, in average, for short-term forecasts, and 8.9% for medium-term forecasts. The results were also supported by applying the Diebold-Mariano test.

A natural extension of this work could be to include exogenous fundamental variables relating with the hourly electricity spot price in order to improve the forecasting accuracy, such as water reservoir levels and load demand patterns. This possibility was not taking into account owing to the difficulty in obtaining some of these data, and because these regressors are available, usually, at low frequencies (e.g., daily, weekly and monthly). This issue is an open direction left for future work. For example, Shang (2013) gives a theoretical possible guidance by considering a functional time series model with discrete and functional regressors. Additionally, a work toward the near future is to compare the results with those obtained by machine learning algorithms such as gradient boosting and its extensions, and time-series models borrowed from Deep Learning using RNN (Recurrent Neural Network) architectures as LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) networks (Lago et al., 2018).

6. ACKNOWLEDGMENTS

The authors thanks Laura Marquéz, M. Sc. student in Economics, for her valuable help with the database. This research was supported by the Grant 13238 from the Comité para el Desarrollo de la Investigación (CODI), University of Antioquia, Colombia.

Appendix I: Functional principal component analysis

The functional PCA (e.g., Dauxois et al., 1982; Hall and Hosseini-Nasab, 2006., Tran, 2008; and Shang, 2014) relies on the spectral analysis of the covariance operator K of *Y(x)* in the class $L_2(\chi)$ of square-integrable functions on χ,

$$K: L_2(\chi) \longrightarrow L_2(\chi)$$

$$\phi \mapsto \mathrm{K}\phi = \int_\chi K(\cdot, v)\phi(v)\,dv$$

where K:$\chi\times\chi\longrightarrow$R is the continuous covariance function of Y(x),

$$K(u,v)=E[Y(u)-\mu(u)(Y(v)-\mu(v))],\ u,v\in\chi$$

By the Mercer's lemma, the spectral decomposition of K is defined as

$$\mathrm{K}\phi_k = \lambda_k\phi_k \to \int_\chi K(u,v)\phi_k(v)\,dv = \lambda_k\phi_k(u), \qquad k = 1,2,\ldots$$

where $\{\phi_k \in L_2(\chi)\}$ is an orthonormal sequence of continuous eigenfunctions, and $\{\lambda_k\}$ the corresponding non-decreasing sequence of non-negative eigenvalues.

The scores of *Y(x)* are $\left\{\beta_k = \int_\chi \left[Y(x) - \mu(x)\right]\phi_k(x)\,dx\right\}$, which are zero-mean uncorrelated random variables with variance λ_k.

Finally, the truncated Karhunen-Loève expansion at the first K terms provides the best approximation of *Y(x)*, given by

$$Y(x) \approx \mu(x) + \sum_{k=1}^{K}\beta_k\phi_k(x) \qquad x \in \chi$$

REFERENCES

Aggarwal, S., Saini, L., Kumar, A. (2009), Short term price forecasting in deregulated electricity markets: A review of statistical models and key issues. International Journal of Energy Sector Management, 3(4), 333-358.

Aneiros, G., Vilar, J., Raña, P. (2016), Short-term forecast of daily curves of electricity demand and price. International Journal of Electrical Power and Energy Systems, 80, 96-108.
Barrientos, J., Rodas, E., Velilla, E., Lopera, M., Villada, F. (2012), Modelo para el pronóstico del precio de la energía eléctrica en Colombia. Lecturas de Economía, 77(2), 91-127.
Barrientos, J., Tabares, E., Velilla, E. (2018), Forecasting electricity price in Colombia. A comparison between neural network, ARMA process and hybrid models. International Journal of Energy Economics and Policy, 8(3), 97-106.
Bello, S., Beltrán, R. (2010), Caracterización y pronóstico del precio spot de la energía eléctrica en Colombia. Revista Maestría en Derecho Económico, 6(6), 293-316.
Besse, P., Cardot, H. (1996), Approximation spline de la prévision d'un processus functionnel autorégressif d'ordre 1. The Canadian Journal of Statistics, 24(4), 467-487.
Conejo, A., Contreras, J., Espinola, R., Plazas, M. (2005), Forecasting electricity prices for a day-ahead pool-based electric energy market. International Journal of Forecasting, 21(3), 435-462.
Congress of the Republic of Colombia. (1994), Laws 142 and 143, July 1994. Available from: https://www.minminas.gov.co.
Crespo, J.C., Hlouskova, J., Kossmeier, J., Obersteiner, M. (2004), Forecasting electricity spot-prices using linear univariate time-series models. Applied Energy, 77(1), 87-106.
Damon, J., Guillas, S. (2015), Far: Modelization for Functional Auto Regressive Processes, R Package Version 0.6-5. Available from: https://www.cran.r-project.org/package=far.
Dauxois, J., Pousse, A., Romain, Y. (1982), Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. Journal of Multivariate Analysis, 12(1), 136-154.
de Livera, A., Hyndman, R., Snyder, R. (2011), Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association, 106(496), 1513-1527.
Diebold, F., Mariano, R. (2002), Comparing predictive accuracy. Journal of Business and Economic Statistics, 20(1), 134-144.
Hall, P., Hosseini-Nasab, H. (2006), On properties of functional principal components analysis. Journal of the Royal Statistical Society: Series B, 68(1), 109-126.
Horváth, L., Kokoszka, P. (2012), Inference for Functional Data with Applications. Vol. 200. New York: Springer Science & Business Media.
Huisman, R., Huurman, C., Mahieu, R. (2007), Hourly electricity prices in day-ahead markets. Energy Economics, 29(2), 240-248.
Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeen, F. (2018), Forecast: Forecasting Functions for Time Series and Linear Models, R Package Version 8.4. Available from: http://www.pkg.robjhyndman.com/forecast.
Hyndman, R., Shang, H. (2009), Forecasting functional time series. Journal of the Korean Statistical Society, 38(3), 199-211.
Hyndman, R., Shang, H. (2010), Rainbow plots, bagplots, and boxplots for functional data. Journal of Computational and Graphical Statistics, 19(1), 29-45.
Hyndman, R., Shang, H. (2017), ftsa: Functional Time Series Analysis, R Package Version 4.8. Available from: http://www.cran.r-project.org/package=rainbow.
Hyndman, R., Ullah, M. (2007), Robust forecasting of mortality and fertility rates: A functional data approach. Computational Statistics and Data Analysis, 51(10), 4942-4956.
Kokoszka, P., Horváth, L. (2017), Introduction to Functional Data Analysis. Texts in Statistical Science. New York: CRS Press.
Lago, J., de Ridder, F., de Schutter, B. (2018), Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Applied Energy, 221, 386-405.
Liebl, D. (2013), Modeling and forecasting electricity spot prices: A functional data perspective. The Annals of Applied Statistics, 7(3), 1562-1592.
Lira, F., Muñoz, C., Nuñez, F., Cipriano, A. (2009), Short-term forecasting of electricity prices in the Colombian electricity market. IET Generation, Transmission and Distribution, 3(11), 980-986.
Panchakshara, G.G., Sedidi, V., Kumar, A., Narayan, B. (2014), Forecasting electricity prices in deregulated wholesale spot electricity market: A review. International Journal of Energy Economics and Policy, 4(1), 32-42.
Portela, J., Muñoz, A., Alonso, E. (2018), Forecasting functional time series with a new Hilbertian ARMAX model: Application to electricity price forecasting. IEEE Transactions on Power Systems, 33(1), 545-556.
R Development Core Team. (2018), R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.r-project.org.
Ramsay, J.O., Dalzell, C.J. (1991), Some tools for functional data analysis. Journal of the Royal Statistical Society: Series B (Methodological), 53, 539-572.
Ramsay, J.O., Silverman, B.W. (2005), Functional Data Analysis. Springer Series in Statistics. 2nd ed. New York: Springer.
Ramsay, J.O., Wickham, H., Graves, S., Hooker, G. (2017), fda: Functional Data Analysis, R Package Version 2.4.7. Available from: https://www.cran.r-project.org/package=fda.
Shahidehpour, M., Yamin, H., Li, Z. (2002), Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management. New York: Wiley.
Shang, H. (2010), Visualizing and Forecasting Functional time Series. PhD Thesis, Monash University, Department of Econometrics and Business Statistics.
Shang, H. (2013), Functional time series approach for forecasting very short-term electricity demand. Journal of Applied Statistics, 40(1), 152-168.
Shang, H. (2014), A survey of functional principal component analysis. AStA Advances in Statistical Analysis, 98(2), 121-142.
Shang, H., Hyndman, R. (2016), Rainbow: Rainbow Plots, Bagplots and Boxplots for Functional Data, R Package Version 3.4. Available from: https://www.cran.r-project.org/package=rainbow.
Taylor, J. (2003), Short-term electricity demand forecasting using double seasonal exponential smoothing. Journal of the Operational Research Society, 54(8), 799-805.
Tran, N.M. (2008), An Introduction to Theoretical Properties of Functional Principal Component Analysis. PhD Thesis, Department of Mathematics and Statistics. Victoria, Australia: University of Melbourne.
Vilar, J., Cao, R., Aneiros, G. (2012), Forecasting next-day electricity demand and price using nonparametric functional methods. International Journal of Electrical Power and Energy Systems, 39(1), 48-55.
Wang, J.L., Chiou, L.M., Müller, R.G. (2016), Review of functional data analysis. Annual Review of Statistics, 3, 257-295.
Weron, R. (2006), Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach. Vol. 403. New York: John Wiley & Sons.
Weron, R. (2014), Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 30(4), 1030-1081.
Weron, R., Misiorek, A. (2008), Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. International Journal of Forecasting, 24(4), 744-763.