Analyzing Electricity Demand in Colombia: A Functional Time Series Approach

Jorge Barrientos Marín[1,2]*, Laura Márquez Marulanda[3], Fernando Villada Duque[4]

[1]Departamento de Economía, Universidad de Antioquia, Colombia, [2]Facultad de Economía, Universidad Autónoma Latino Americana - UNAULA, Colombia, [3]Escuela de Ciencias Económicas y Administrativas, Universidad EIA, Colombia, [4]Departamento de Ingeniería Eléctrica, Universidad de Antioquia, Colombia. *Email: jorge.barrientos@udea.edu.oc

ABSTRACT

In this work we are interested in analyzing the energy demand in Colombia for a short-term horizon, from a functional data approach. First, we make an exhaustive review of the literature on functional spaces as a potential source of statistical information. It is, of course, a theoretical reinterpretation since in practice the data are elements of a finite-dimensional space; however, very high-frequency data, properly treated, can be viewed as elements of a space of continuous functions. Second, we put such a reinterpretation into practice, by performing a spline-type smoothing of commercial energy demand, based on hourly-daily data. As a result, a function or smooth curve is obtained for each day. Finally, we expose the usefulness of this new approach for statistical analysis, modeling, and projection (or forecasting) of stochastic processes that generate high-frequency random variables.

Keywords: Functional Data, Functional Time Series, Data Smoothing, Energy Demand
JEL Classifications: C32, C50, C55, L94, Q41, Q47

1. INTRODUCTION

In the mid-1990s, a significant number of countries began to deregulate their electricity industries and to restructure their energy markets. There was an assortment of motivations: inefficiencies in electricity pricing; improving quality and expanding service coverage; diversifying the energy matrix; and guaranteeing energy security which, for example, in many countries such as Colombia, was low due to both strikes by workers in thermoelectric plants and extreme weather events –including El Niño– which in 1992 forced the national government to strictly ration energy consumption for 11 months.

Colombia was not the exception. Such deregulation emerged with laws 142 and 143 of 1994 (Congreso de la República, 1994a, b) through which a legal framework was established to make it effective. Specifically, it created the energy exchange in 1995

in addition to the institutional framework required for market operation and the regulation, monitoring, and sanctioning of agents from the entire energy chain –which must operate in the national territory under a regulated-competition scheme with participation of public and private companies. The significance of this market lies in that energy is a commodity that, unlike other types of commodities, can only be produced at a generation plant if an effective demand exists almost concurrently, a characteristic that makes the trading and exchange process of energy production distinct with respect to other services or public utility goods.

In this paper we focused on the wholesale energy market in Colombia, hence it is important to define how transactions and the determination of variables in it. Energy transactions are determined in two ways: short and long-term transactions. In the long-term market, trading agents and generators file their

energy sale-purchase contracts with the Administrator of the Commercial Exchange System -ASIC- (Administrador del Sistema de Intercambios Comerciales, for its name in Spanish), who will determine the hourly transactions for each agent. On the other hand, there is a short-term market that operates daily with hourly resolution; demand and price are therefore determined through an auction system. In Colombia, there is a daily energy market with hourly resolution; therefore, both demand and price are determined through an auction system. This is a suitable scenario for the application of functional methodology given the wide availability of data and how frequently they are represented in functions, by expressing each day as a function composed of 24-hourly observations.

This work aims to apply a Functional Data Analysis -FDA- method to a time series, specifically electricity demand time series. Given the high frequency of data and data behavior, FDA offers advantages in short-term statistical analysis, as well as projections of variables, over traditional methodologies based on finite-dimensional vector processes. First, it allows for the numerical representation of infinite dimensional objects; that is, continuous-time functions are accepted as representations both parametrically, and more importantly, nonparametrically. Second, it reduces measurement noise. Third, it makes it possible to analyze data taken at any given time, especially if time observations are not equally spaced. Fourth, traditional statistical methods with functional data fail due to the ratio between the number of variables and the sample size (each variable is a discretized point) and the strong correlation among the variables that becomes an issue as their conditioning is problematic in the context of multivariate models.

We emphasize that it is necessary to understand an unusual market with unique characteristics for any country where electricity is provided under regulated competition; electricity is a public utility good with an inelastic demand in the short term since, in the regulated market, consumers' decisions are not always price based. Moreover, it should be considered that the magnitude of daily consumption is the result of the energy consumption habits by households and economic activities.

In Colombia, there are very few studies that provide electricity demand statistical analysis by using intra-hour data, which is relevant for a market operating daily with hourly resolution, like the Colombian one. Accordingly, the contribution of this work may be summarized as follows: first, it enables the analysis of demand characteristics for each day capturing hourly information, achieved by expressing data as time functions. Second, the proposed modelling method offers a short-term projection for energy demand that includes a novel treatment of atypical data attributed to external factors associated with regulatory or climatic factors. Third, to the best of our knowledge, this is the first study nationwide to use an FDA-based approach to analyze energy demand evolution in Colombia, thus making an important contribution to the national literature. Finally, this work paves the way for future research, expanding the spectrum of application to other market variables such as prices and hydrology, to name a few, and even to other economic sectors where similar information is generated.

This research work is divided into seven sections: first one, this introduction. Second, we point out the importance of analyze electricity demand in Colombia; third, we carry out a literature review, including the theoretical framework synthesizing some functional methodologies as well as some of the traditional more employed to study energy market variables. Forth, the functional methodology and data description are outlined. Fifth, the smoothness procedure and the result of the modelling is discussed. Sixth, we discuss what can be done once we get curves as a data. Finally, the conclusions are presented.
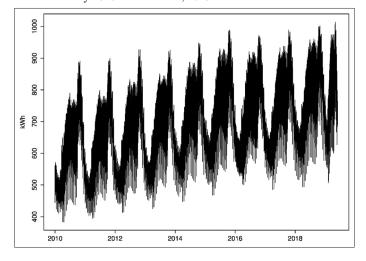
2. WHY WE STUDY ELECTRICITY DEMAND

The need and usefulness to carry out a statistical analyzing of the evolution over time of the electricity demand are justified for several reasons. First, it is a crucial input for large-scale industrial activity. Second, energy demand is a key predictor of economic activity, most notably manufacturing industry. Third, forecasting energy demand in the medium term helps determine whether it is worth expanding the installed generation capacity. Fourth, growing demand calls for energetic security and efficiency to be guaranteed; it is therefore an indicator of the degree to which the energetic matrix should de diversified. For example, thermal generation accounts for approximately 30 % of the Colombian electricity system and the rest is generated through hydroelectric power, meaning there is a risk of market shortages primarily due to climatic factors. Lastly, demand is a fundamental variable in price formation and evolution on the stock market.

Proper energy demand statistical analysis and its projection, would contribute to anticipating scenarios to make timely decisions, reducing the likelihood of all end users being affected. It would also improve the reliability of service provision as well as trust in all the agents involved in the sector. Thus, the overall aim of this work is to study the dynamics and evolution of energy demand through the application of statistical and econometric tools that provide accurate modelling, producing reliable information which supports sound decision-making and guarantees the operation of the electricity market and the economy at large.

Trader firms represent the entire market demand. Because of auction to provides electricity in the market is a kind of closed seal, which every generator offers only one price and one specific quantity of electricity for each hour of the day, the demand is also determined on an hourly basis and is so recorded on the daily operations, this information is report by the company XM (www.xm.com.co), the wholesale energy market operator in Colombia. As usual features the energy demand in Colombia shows significant changes across hours, days, and weeks; however, its behavior is stable in the long term, showing an increasing trend and strong seasonality, as we can see in Figure 1 which shows electricity demand from January 2010 to May 2019.

Figure 1: Hourly energy demand in Colombia from January 2010 to May 2020. Source: XM, 2019. Own calculations

3. LITERATURE REVIEW AND METHODOLOGIES FOR ANALYZING TIME SERIES

Several methodologies have been proposed in the literature to study electricity demand and make statistical inference or generate reliable energy demand projections, they are focused on different aspects such as variability, level changes, determinant factors, household and industry consumption evolution, and the regulations affecting this market. Many of these studies provide policy recommendations and regulatory modifications, for instance, by creating long-term contracts to keep the price at adequate levels, by guaranteeing the coverage of demand or by increasing the supply, encouraging new generators to enter the market. In general, through decisions that maintain the stability in energy sector.

In more conventional literature, there are several authors who use time series methods to model and forecast energy price and demand with different time horizons; it is also possible to identify models that seek inclusion of exogenous variables employing smoothing filters, for example, exponential. Time series application includes autoregressive integrated moving average processes, long memory process, seasonality, threshold, and conditional heteroscedasticity models. Barrientos et al. (2007) propose nonparametric regression models to modelling the daily demand of electric energy in southwestern Colombia, considering influencing factors such as times, the day of the week analyzed, month, and year. Botero and Cano (2008) highlight the need for measuring and modeling several types of variables in the Colombian energy sector that include the restructuring it has experienced over the years, since electric energy variables (demand and prices) are characterized by having great volatility, representing a risk for the energy market and the agents.

In this line, traditional models are implemented with some variations for market adaptation, such as He and Lin's (2018) present a study in which an ADL-MIDAS (autoregressive distributed lag mixed data sampling) model is built to identify the optimal model among combinations of different weight

functions and forecasting methods to predict energy demand in China. Yuan et al. (2017) use probabilistic models in different scenarios to identify characteristics and changes in energy demand at the regional and national levels. They establish some influence factors as variables that could modify or influence energy demand, determining this effect by using a hierarchical Bayesian model. The authors conclude that this Bayesian model structure incorporates fundamental information for improving the estimation, provided that the effects' behavior is known.

Saloux and Candando (2018) predict the heating demand, obtained from different algorithms, by comparing two different scenarios: one with actual weather conditions supported by data, and the other using weather forecasts. The authors employ learning techniques and models such as neural networks and vector support machines, highlighting the usefulness of having training and validation samples for algorithm learning. Chou and Tran (2018) provide an interesting application forecasting energy consumption using machine learning techniques based on usage patterns of residential householders.

Functional data analysis is relatively new topic in econometrics due to the limited availability of high-frequency data and the fact that the information-gathering technologies required have only been recently developed, as well the econometric models to approach this type of analysis. Functional data are found in multiple fields, including spectrometry (data for the analysis of chemical substances); radio frequency measured in kHz (voice recognition data); and, naturally, in electricity prices and demand (measured in hours, days, or weeks). However, their application to other disciplines is still limited. The most used methods to model functional data are parametric methods such as functional linear regression; nonparametric methods, including functional principal component analysis (FPCA); methods based on Kernel estimators; and base expansion methods (Fouries bases, B-spline) to smooth curves, whose main challenge is the estimation of functional data from noisy observations or, in other words, discrete observations. A popularized method for forecasting is the decomposition of smoothed curves into their estimated functional principal components,

Two classic books on functional data are due to Ramsay and Silverman (1997, 2002, 2005), in which they provide the basic philosophy of functional data analysis, they suggest that researchers must think of observed data functions as single entities, rather than as sequences of individual observations. Moreover, they express that "the term functional in reference to observed data refers to the intrinsic structure of the data rather than to their explicit form". Of course, as the authors say, in practice "functional data are observed discretely as pairs (t_j, x_j) where x_j is a snapshot of the function of time t_j".

After an exhaustive literature review, we did not find as many papers on functional data analysis as we expected. Among the most interesting paper we found: Ramsay and Dazell (1991), one of the first papers on functional data, they provide some tools for functional data analysis; Cardot et al. (1999) develops a functional linear models; Bosq (2000) which develops

properties of linear process in functional spaces; Cardot et al. (2003), they use functional approach for predicting use land; in a very interesting papers, due to Cardot et al. (2003), a testing hypothesis procedure for functional linear model is fully developed; in and extension of this latter paper, Cardot et al. (2003) estimate functional linear models by performing Spline methodology; Cardot et al. (2004) perform testing no effect in functional linear models; Fan and Li (2004) develops statistical test for significance when data are curves; Ferrty and Vieu (2002) carry out an interesting application modelling spectrometric data by performing functional non-parametric regression; Ferrty and Vieu (2003) use non-parametric functional approach for curve discrimination; Ferraty and Vieu (2004) develop nonparametric models for functional data with applications in time series; Ferraty et al. (2006) provide estimation of some characteristics of the conditional distribution in non-parametric functional models. Hyndman and Ullah (2007) propose a new method for robust forecasting of age-specific mortality and fertility rates by performing functional data models. Hyndman and Shang (2009) suggest that using a nonparametric smoothing technique helps to reduce observation error, removing effects of little relevance and solving inverse and multicollinearity problems caused by the amount of data used. Barrientos et al. (2010) develops a new local linear density estimator in kernel regression models for functional data.

Ullah and Finch (2103) carried out a review of the literature on FDA, they found that in total 84 FDA applied articles were identified and 75% has been published since 2005, which coincided with the ending period when most of the theoretical paper were published. They also found that biomedicine was the field with more published applied papers. Clearly, the FDA theory had advanced faster than the technique/technology to collect statistical data that could be interpreted as functions or curves.

For electricity markets few papers, applying functional data tools, was identified. We just mention five of them. Vilar et al. (2012) forecast the next-day demand and price by performing non-parametric functional methods, specifically using functional explanatory data and semi-functional partial linear model, for electricity market of mainland Spain (2008-2009). Liebl (2013) puts forward the perspective of functional data and points out the drawbacks of time series and the traditional methodologies that use them, as issues emerge when replicating fluctuations in spot energy prices. It also emphasizes the significance of the functional relationship between spot prices and energy demand, shows how this relationship is modeled by daily functions, and uses a functional factor model by parameterizing the functions.

Shang (2013) presents an application of functional methodology for the energy market with half-an-hour frequency variables. The author's previous works were conducted with demographic data that replicated the same frequency, for example, for fertility rates. The author argues the usefulness of functional methods lies in that they allow the division of a seasonal univariate time series into a curve time series, that is, functions; in this way, dimensionality is reduced, and the application of functional principal component analysis is rendered possible; the prediction of univariate time series and regression techniques takes place afterward. This work was developed using data of half-hourly, Monday-to-Sunday electricity demand in South Australia.

Aneiros et al. (2016) provide two methods to predict next-day electricity demand and price daily curves given information from past curves. They are based on using robust functional principal component analysis and nonparametric models with functional both response and covariate. Gallón and Barrientos (2021) use functional data analysis and B-Spline joint to Functional Component Principal Analysis to forecast the price of kWh generation in the electricity spot market, since it is a set of appropriate methodologies for modeling and prediction within a very short-term horizon; and in the long term errors may arise due to the existence of atypical data vectors, irregular measurements and the presence of structural alterations resulting from the actions by surveillance and regulatory authorities.

4. FUNCTIONAL DATA ANALYSIS

High frequency data are yielded when determining daily and hourly prices, supply, and demand, which constitutes a major advantage over other types of data structures. They can be interpreted as daily time curves/functions pertaining to a theoretically infinite dimensional space –or functional time series data– as they are known in the literature, although the term "functional" refers to its structure rather than to its form. Two important notions are needed for understanding the concepts: First, a random variable X is called functional if it takes values in a functional space. An observation χ of X is called a functional data. It is worth noting that when X (respectively χ) denotes a random curve, it is implicitly asserted that $X = \{X(t):t \in T\}$ and that $\chi = \{\chi(t):t \in T\}$ respectively. Second, a set of functional data $\chi_1, \chi_2,..., \chi_n$ is the observation of functional variables $X_1, X_2,...,X_n$. This definition spans many scenarios, the most common being curves dataset.

FDA assumes that the sample of random variables of hourly energy demand is a set of infinite dimension functions. The observations of the variable are analyzed as the path of a process where the values for each hour of the day form a discrete set of observed data; the set of observations at a given moment corresponds to a discrete data clustering and a function such that $\chi = \{\chi(t):t \in \mathcal{T} \subset \mathbb{R}\} \in \mathcal{F}$ defined in a set $\mathcal{T}$. The exact form for viewing functions as discrete data sets is as follow: let

$$\left\{\chi_i = \{\chi_i(t); t \in (t_{min}, t_{max})\}\right\}_{i=1,...,n}$$ be a functional dataset and consider its discretized version

$$\left\{\chi_i = \{\chi_i(t_1), \chi_i(t_2),...,\chi_i(t_J)\}\right\}_{i=1,...,n}$$ which can be viewed as a classical matrix. So, the following array:

$$\begin{bmatrix} \chi_1(t_1) & \chi_1(t_2) & \cdots & \chi_1(t_J) \\ \chi_2(t_1) & \chi_2(t_2) & \cdots & \chi_2(t_J) \\ \vdots & \vdots & \vdots & \vdots \\ \chi_n(t_1) & \chi_n(t_2) & \cdots & \chi_n(t_J) \end{bmatrix}$$

can be viewed as J measurements at $t_1, t_2, …, t_J$ of n observed curves. This kind of matrixial data representation is the base for the empirical analysis of theoretical functional data. These estimated functions have a problem known in the literature as the curse of dimensionality, typical of functional time series, for which two solution strategies exist: nonparametric smoothing and dimension reduction.

In this paper we obtain the curves by performing non and semiparametric smoothing Spline methods and principal components analysis, applied for each data set corresponding to the days of the week. The projections for electricity demand are carried out for each day of the week since demand behavior presents a greater degree of similarity depending on the day of the week than between hours from different days. This is explained mainly because economic activities are mostly carried out on weekdays, whereas energy demand on weekends is lower throughout the day. Figure 2 illustrates the hourly demand behavior taking as example the days Tuesdays, from January 2000 to May 2019. In this figure, the strong seasonality of demand can be observed, although it becomes stable around the mean.
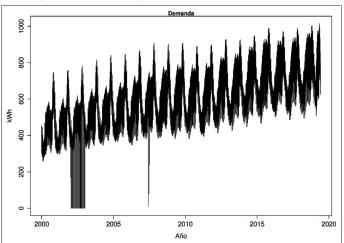
5. EMPIRICAL STRATEGY FOR SMOOTHING ELECTRICITY DEMAND

As previously explained, the methodology described below employs data from the total commercial energy demand expressed in kilowatts/hour, between January 2000 and May 2019; the data set thus corresponds to $\{D_\tau, \tau \in [1, T]\}$, T=170184. Once this time series was formed, it is divided into daily data sets, where T_i denotes the number of points of time data for each i–th day taking the set of information that transforms the energy demand data for different hours of the day into a set of discrete data and later into functions.

$$\{D_{i,\tau}, \tau \in [1, T_i]\} \quad i = \text{Monday}, …, \text{Sunday}$$

Hourly time series curves are obtained by converting the time points T_i into daily functions. These functions $y_{i,t}(x)$ are determined

by $P_{i,T}$, energy demand time series being hourly and $J = 24$, the maximum of periods for each observation.

$$\mathcal{Y}_{i,t}(x) = \{P_{i,\tau}, \tau \in [J(t-1), Jt]\}, \ t = 1, …, n_i, \ \text{for } i = \text{Mon}, …,$$
Sun
$$(1)$$

It is assumed that the corresponding observations are discretizations generated when evaluating a set of unknown smooth functions $f_{i,t}(x)$ at moment x_j, i.e., $y_{i,t}(x_j) = f_{i,t}(x_j)$ for $j=1, …, J$. Generally, it is also assumed that the observational error term is a process data sequence $\varepsilon_{i,t,j}$ such that $y_{i,t}(x_j)$ satisfies the following functional regression model:
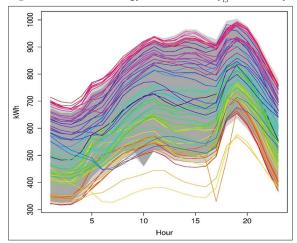
$$\mathcal{Y}_{i,t}(x_j) = f_{i,t}(x_j) + \sigma_{i,t}(x_j)\varepsilon_{i,t,j}, \ t = 1, …, n_i, \ j = 1, …, J,) \quad (2)$$

where $\varepsilon_{i,t,j}$ are independent and identically distributed, $\mathbb{E}(\epsilon_{i,t,j}^2) = 1$, and $\sigma_{i,t}(x_j) > 0$ for all i and t that can vary with x.

When there are errors, the sample of unknown smooth functions $f_{i,t}(x)$ for each i and t can be estimated from the pairs observed $\{(x_j, \mathcal{Y}_{i,t}(x_j))\}_{j=1}^J$ applying nonparametric curve estimation methods. In this case, the smooth curves $\hat{f}_{i,t}$ are estimated with B-Splines. This method consists in the generation of some smoothed functions from certain points, in this case, hourly data, and some basic parametric smoothing points. The foregoing is carried out using the ftsa of R as Hyndman and Shang (2017), where the regulation parameters or degree of smoothing λ_i are chosen through generalized cross-validation. This method consists in choosing the optimal smoothing parameter based on error minimization with respect to a data sample; hence, the process is repeated until an optimal λ_i is obtained.

The smoothing of the functions is required because data such as hourly observations do not allow to easily model energy demand and its oscillations throughout the day; in a functional manner, the behavior of the interest variable is better visualized over time. This step also facilitates identifying those functions made up of observations that correspond to atypical data, since they do not follow the standard behavior of a given day, moving away from

Figure 2: Hourly energy demand in Colombia for Tuesdays

Figure 3: Smoothed energy demand curves $\hat{f}_{i,t}$ for Tuesdays

the variable's habitual trend. Due to the high dimensionality of the functional data, it is common that some curves be itself extreme values or have a shape very different from the rest of the curves in the sample.

Figure 3 illustrates the smoothed curves of energy demand for Tuesdays in the period analyzed; additionally, some of the atypical curves can be observed. Considering the presence of atypical data, it is relevant to show some statistics of the data presented in Table 1, where the number of observations (per hour), the number of days, the minimum and maximum demand values for the mentioned day, as well as the mean and median. From these descriptive statistics, it can be validated that weekdays are those with greater energy demand, whereas weekends show lower values, possibly associated with the level of economic activity characteristic of weekdays.

The identification and elimination of functional atypical data help improve the accuracy of principal components since they can cause erroneous conclusions. To detect functional atypical values, Hyndman and Shang (2010) propose a bivariate and functional boxplots of High-Density Region (HDR) defined as $R_\alpha = \left\{ z : \hat{f}(z) \geq f_\alpha \right\}$, where $\hat{f}(z)$ is the estimated bivariate density function of the first principal components of the smoothed functions; the points inside R_α have greater density than those outside the region.

In Figure 4 the diagram showing the zone corresponding to the greatest density can be observed.[1] It is possible to observe the bivariate (left) and functional (right) HDR boxplots of electricity demand 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. Atypical values,

---

1   This diagram was generated by using the package ftsa in R, developed by the R Development Core Team (2010)

Table 1: Descriptive statistics of energy demand – With atypical data

Day	T_i	n_i	Min	Mean	Med	Max
Monday	24312	1013	2576948	6416358	6357520	10133466
Tuesday	24312	1013	56	6647886	6640183	10142301
Wednesday	24312	1013	3163948	6680957	6660578	10022255
Thursday	24336	1014	121228	6665021	6637431	10105364
Friday	24312	1013	1332015	6643439	6627556	9992638
Saturday	24312	1013	2534978	6261134	6275393	9535581
Sunday	24288	1012	2774289	5642098	5586426	8965678

Table 2: Descriptive statistics of energy demand – Without atypical data

Day	#outliers	n_i	Min.	Mean	Med.	Max.
Monday	53	964	2915291	6441722	6391486	9997495
Tuesday	49	964	3035384	6689646	6677931	10142301
Wednesday	49	964	3178525	6693993	6675023	10022255
Thursday	50	964	3179461	6712089	6694777	10105364
Friday	50	963	1332015	6679779	6674468	9992638
Saturday	50	964	2534978	6273126	6288476	9535581
Sunday	49	964	2923257	5650728	5592284	8965678

in this case, correspond to dates between mid-2015 and early 2016; these curves are part of the El Niño phenomenon time, a climate change phenomenon of low precipitations and heat waves increasing energy demand. Table 2 and Figure 5 show the statistics of the data and the smoothed curves of electricity demand without atypical values.

To be able to capture the nature of electricity demand fluctuations, an approach of functional principal components is applied as in Hyndman and Shang (2009). The principal components approach seeks to transform a set of variables into another one in which these are not correlated and keep their variance and, consequently, the most important information contained in them. The approach of principal components extended to functional data implies generating a functional time series which is decomposed into a linear combination of numbers of proper functions from a function of the expression for each day i. Equation 2 shows the decomposition of the functions into their proper functions:

$$f_{i,t}(x) = \mu_i(x) + \sum_{k=1}^{K_i} \beta_{i,t,k}\phi_{i.k}(x) + \xi_{i.t}(x), \ K_i < n_i, \ (3)$$

where $\mu_i(x)$ is the mean of the function i, $\left\{ \phi_{i,k}(x) \right\}_{k=1}^{K_i}$ is the set of orthonormal proper functions or principal components with the corresponding coefficient (scores) variable over time $\left\{ \beta_{i,t,k} \right\}_{k=1}^{K_i}$, K_i is the number of components per day, and $\left\{ \xi_{i,t}(x) \right\}_{t=1}^{n_i}$ is centered on random functions. An estimator for $\mu_i(x)$ is obtained using the method proposed by Hyndman and Shang (2009). By estimating $\left\{ \hat{\phi}(x) \right\}_{k=1}^{K_i}$ consistent for $\left\{ \phi(x) \right\}_{k=1}^{K_i}$, the proposed method is obtained implementing R's FTSA library.

Figure 6 shows the functional principal components in the electricity demand curves corresponding to Tuesdays, employing an ARIMA model for the forecast. Based on the estimated empirical mean functions $\left\{ \hat{\mu}_i(x) \right\}$, functional principal components $\left\{ \hat{\phi}_{i,k}(x) \right\}$, and the corresponding dynamic coefficients $\left\{ \hat{\beta}_{i,t,k} \right\}$, we obtain:

$$\mathcal{Y}_{i,t}(x_j) = \hat{\mu}_i(x) + \sum_{k=1}^{K_i} \hat{\beta}_{i,t,k}\hat{\phi}_{i,k}(x_j) + \hat{\xi}_{i,t}(x_j) + \hat{\varepsilon}_{i,t,j},$$
$$t = 1,\ldots,n_i, \ j = 1,\ldots,J \tag{4}$$

6. WHAT CAN BE DONE WITH FUNCTIONS: A NAÏVE EXAMPLE

Due to the orthonormality property of the basis functions, each one of the coefficients $\left\{ \hat{\beta}_{i,t,k} \right\}_{k=1}^{K_i}$ can be forecast independently, for example, with an ARIMA-type model. Conditional to the observed data $D_i = \{y_{i,t}(x_j) : t = 1,\ldots,n_i, j = 1,\ldots,J\}$, and the estimated principal components $\Phi_i = \left\{ \hat{\phi}_{i,1}(x),\ldots,\hat{\phi}_{i,K_i}(x) \right\}$ for each day, an

Figure 4: HDR bivariate and functional boxplots of energy demand for Tuesdays. The numbers in HDR bivariate boxplot correspond to atypical days
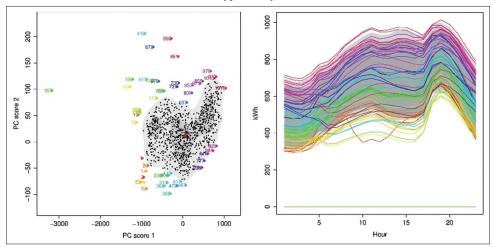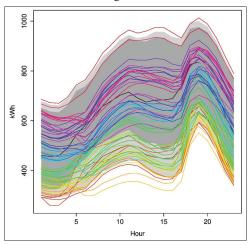

Figure 5: Smoothed energy demand curves for Tuesday $\hat{f}_{i,t}$ after removing the outliers

h-step-ahead forecast of $\mathcal{Y}_{i,n_i+h}(x)$, $h \in \mathbb{N}$, is obtained by performing the forecast rule based on the conditional expectation that minimizes the MSE of $y_{i,t}(x_j)$:

$$\hat{\mathcal{Y}}_{i,n_i+h|n_i}(x) = \mathbb{E}\left[\mathcal{Y}_{i,n_i+h}(x) \mid D_i, \Phi_i\right] = \hat{\mu}_i(x) + \sum_{k=1}^{K_i} \hat{\beta}_{i,n_i+h|n_i,k}\hat{\phi}_{i,k}(x)$$

i= Mon,…, Sun

where $\hat{\beta}_{i,n_i+h|n_i,k}$ defines the h-step-ahead forecast of $\beta_{n_i+h.k}$ for day i.

The optimal number of principal components was chosen using the method of Shang (2013) for which the sample of curves for each day is divided into a training set with $n_i^*=n_i-l$ functions, and a validation set with $l=52$ curves corresponding to the weeks of the year. Next, a forecast accuracy measure is obtained for the functions in the validation set based on the functional time series model adjusted with the training set based on the forecast accuracy. Forecast accuracy through the mean absolute percent error (MAPE) of $\left(\left|p_{h,j}\right|\right)$ can be observed in Tables 3 and 4, which is defined as follow:

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

Table 3 reports the optimal number of K_i, note that it is quite variable from Monday to Sunday: for Mondays, Wednesdays and Fridays the optimal number of K_i is 8; for Saturday it is 10; for Tuesdays and Thursdays it is 4; and for Sundays it is 2. The last row in the next table indicates the variation proportion which is explained by the optimal principal components K_i, whose value for all the days is around 99%. Similar results are obtained with the root mean square percentage error (RMSPE), which is defined as the deviation function of the square root of the sum of square errors for a data set or time series.

The mean absolute percentage error (MAPE) is a statistical measure of a forecast system's accuracy widely used in the literature since it measures this accuracy as a percentage and can be calculated as the average absolute percentage error for each period minus the real values divided by the real values. In this case, it is found that the first term of the numerator corresponds to the real value, whereas the second term corresponds to the forecast value. In turn, the root mean square error (RMSPE) is also used, providing the same results as MAPE.

After obtaining the optimal number of proper functions and the corresponding empirical principal components $\left\{\hat{\phi}_{i,k}(x)\right\}_{k=1}^{K_i}$, and by using the ARIMA time series model, 48 h of forecast per day are obtained. This section, that just pretend illustrate with a simple example, is centered on the day-ahead projections *(h=1)*, concluding that the model works well in to project the structure of these functions.
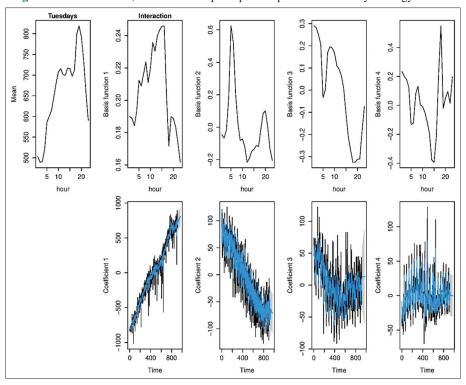
From the graphs it is observed that forecasts for each day of the week capture the common structural patterns that characterize the variability of hourly electricity demand. That is to say, the shape of the forecast curves shows that from around 6:00 a.m., the demands tend to increase progressively until 12:00 noon, it declines until 3:00 p.m. when it then increases until reaching

Table 3: Mean Absolute Percentage Error (MAPE) - ARIMA

K	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	0.5820	0.3991	0.2606	0.3473	0.2675	0.2774	0.2928
2	0.5823	0.3981	0.2631	0.3465	0.2699	0.2787	0.2731
3	0.5826	0.3994	0.2639	0.3471	0.2697	0.2789	0.2930
4	0.5827	0.3911	0.2649	0.3461	0.2701	0.2801	0.2938
5	0.5825	0.4011	0.2651	0.3501	0.2712	0.2800	0.2991
6	0.5828	0.3989	0.2514	0.3469	0.2695	0.2822	0.2954
7	0.5829	0.3987	0.2552	0.3489	0.2700	0.2813	0.2987
8	0.5818	0.3962	0.2491	0.3466	0.2686	0.2886	0.2971
9	0.5822	0.3971	0.2519	0.3511	0.2790	0.2884	0.2999
10	0.5821	0.3983	0.2586	0.3490	0.2777	0.2710	0.2901
Prop.	0.9898	0.9988	0.9999	0.9989	0.9998	0.9799	0.9959

Prop: Variation proportion explained by the optimal principal component

Figure 6: Mean function, first functional principal component for Tuesday's energy demand

a second peak around 8:00 p.m., and then it starts to decline progressively.

The forecast was compared with an ARFIMA model. There are several reasons for using an ARFIMA model and apply it to the smoothed data. First, data are of high frequency (daily with hourly resolution); second, a quick adjustment of the data yields a fractional root close to 0.5; third, the above implies carrying out a long memory contrast, which was done using the contrast of Qu (2011), but it rejects long memory H0 in favor of spurious long memory (it yields some critical values below 1.7, at any significance level); fourth, using a contrast of Hylleberg et al. (1990) the H_0 of stationary unit root is not rejected. Thus, an ARFIMA seems appropriate, at least as a comparison model. Figure 7 illustrates the functional principal components in the energy demand curves corresponding to Tuesdays.

The optimal number of principal components is presented in Table 4, based on the MAPE. As ARIMA model, in the ARFIMA case the

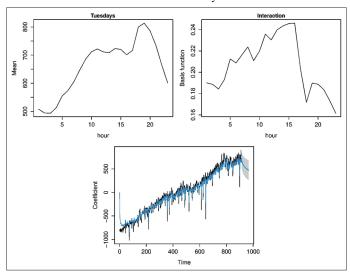**Figure 7:** Mean function, first functional principal component and score for tuesday
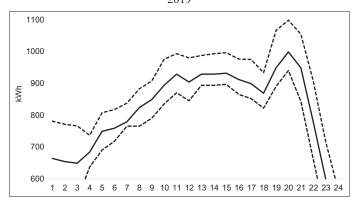
Table 4: MAPE- ARFIMA

K	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	0.5840	0.3920	0.2665	0.3470	0.2695	0.2789	0.3041
2	0.5844	0.3931	0.2670	0.3465	0.2698	0.2799	0.3054
3	0.5870	0.3940	0.2673	0.3473	0.2769	0.2799	0.3030
4	0.5868	0.3911	0.2689	0.3461	0.2770	0.2801	0.3038
5	0.5900	0.4001	0.2641	0.3501	0.2771	0.2800	0.3091
6	0.5802	0.4015	0.2664	0.3469	0.2769	0.2822	0.3042
7	0.5909	0.4019	0.2692	0.3489	0.2770	0.2813	0.3087
8	0.5818	0.4028	0.2606	0.3466	0.2686	0.2886	0.3101
9	0.5949	0.4028	0.2701	0.3511	0.2778	0.2884	0.3110
10	0.5951	0.4103	0.2698	0.3494	0.2779	0.2774	0.2928
Prop.	0.9998	0.9988	0.9998	0.9989	0.9998	0.9999	0.9959

Prop: Variation proportion explained by the optimal principal component, MAPE: Mean absolute percentage error

Figure 8: Projected electricity demand on first Tuesday for January 2019

optimal number of K_i is also quite variable from Monday to Sunday: for Monday, Wednesday and Friday, the optimal number of K_i is 8; for Saturday it is 10; for Tuesday and Thursday it is 4; and for Sunday it is 2, which is the same results in terms of number of optimal components. Of course, very similar results are obtained with the RMSPE.

The Figure 8 shows the FTS point projection (forecast) for the first Tuesday for January 2019 withs its respective 95% bootstrap confidence bands; we can observe this projection exhibit (or capture) the common structural features of the hourly electricity demand described in Section 2. This projection shows exactly the intra-day patter jointly the familiar picks at the hours of higher electricity consumption. It is worth noting to show that MAPE for FTA forecasting is slightly smaller than MAPE for ARFIMA forecasting, so in this case FTA provided slightly better forecasting results when the performance is evaluated with MAPE but is no conclusive if RMSPE is applied.

7. CONCLUSION

This paper applied a methodology for the analysis of functional time series (FDA) to electricity demand in Colombia. Due to the high frequency of the data and their evolution over time, FDA presents advantages in the short-term analysis over traditional methodologies based on vector processes of finite dimension. Specially in capture hidden patterns related to intra-day and intra-hour consumer's behavior.

A more sophisticated modelling and forecasting of the demand could be carried out by including: a) seasonality, due to strong seasonal effect in the demand evolution over times as showed in the Figure 1; b) it is necessary to estimate several models in order to benchmark the FTS model, as SARIMA, Holt-Winter models, Neural Networks (NNAR) or vectorial support machine; c) clearly, this research can be extended by using other functional nonparametric methodologies, as liner functional model as in Bosq (2000) or based on kernels as in Ferraty and Vieu (2006) and including fundamental variables such as price per kWh, river water volumes or hydric reserves; as well as variables indicating economic activity, GDP or industrial production indices, although these indicators of the economic activity present difficulties due to its low frequency measurement.

8. ACKNOWLEDGEMENT

REFERENCES

Aneiros, G., Vilar, J., Raña, P. (2016), Short-term forecast of daily curves of electricity demand and price. Electrical Power and Energy System, 80, 96-108.

Barrientos, A.F., Olaya, J., González, M.V. (2007), Un modelo spline para el pronóstico de la demanda de energía eléctrica. Revista Colombiana de Estadística, 30(2), 187-202.

Barrientos-Marín, J., Ferraty, F., Vieu, P. (2010), Locally modelled regression and functional data. Journal of Nonparametric Statistics, 22(5), 617-632.

Bosq, D. (2000), Lecture Notes in Statistics: Linear Process in Functional Spaces, Theory and Applications. Vol. 149. New York: Springer Verlag.

Botero, S., Cano, J.A. (2008), Análisis de series de tiempo para la predicción de los precios de la energía en la bolsa de Colombia. Cuadernos de Economía, 27(48), 173-208.

Cardot, H., Faivre, R., Goulard, M. (2003), Functional approaches for predicting land use with temporal evolution of coarse resolution remote sensing data. Journal of Applied Statitics, 30, 1185-1119.

Cardot, H., Ferraty, F., Mas, A., Sarda, P. (2003), Testing hypothesis in the functional linear model. Scandinavian Journal of Statistics, 30, 241-255.

Cardot, H., Ferraty, F., Mas, A., Sarda, P. (2004), Spline estimators for the functional linear models. Statistica Sinica, 13, 571-591.

Cardot, H., Ferraty, F., Sarda, P. (1999), Linear functional models. Statistics and Probability Letters. 45, 11-22.

Cardot, H., Goia, A., Sarda, P. (2004), Testing for no effects in functional linear regression models, some computational approaches. Communications in Statistics, Simulation and Computation, 33, 179-199.

Chou, J.S., Tran, D.S. (2018), Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy, 165, 709-726.

Congreso de Colombia a. (1994), Ley 143 de 1994. Bogotá: Diario Oficial No. 41.434.

Congreso de Colombia b. (1994), Ley 142 de 1994. Bogotá: Diario Oficial No. 41.433.

Fan, J., Li, S. (2004), Test of significance when de data are curves. Journal of American Statistical Association. 93, 1007-1021.

Ferraty, F., Laksaci, A., Vieu, P (2006), Estimation of some characteristics of the conditional distribution in non-parametric functional models. Statistical Inference for Stochastic Processes, 9, 47-76.

Ferraty, F., Vieu, P. (2006), Nonparametric Functional Data Analysis: Theory and Practice. Germany: Springer Science and Business Media.

Ferrty, F., Vieu, P. (2002), The functional nonparametric model and application to spectrometric data. Computational Statistics, 17, 545-564.

Ferrty, F., Vieu, P. (2003), Curves discrimination: A nonparametric functional approach. Computational Statistics and Data Analysis, 44, 161-173.

Ferrty, F., Vieu, P. (2004), Nonparametric models for functional data with applications in regression, times series prediction and curves discrimination. Journal of Nonparamtric Statistics, 16, 111-127.

Gallón, S., Barrientos-Marín, J. (2021) Forecasting the Colombian electricity spot price under a functional approach. International Journal of Energy Economics and Policy, 11(2), 67-74.

He, L., Lin, B. (2018), Forecasting China's total energy demand and its structure using ADL-MIDAS model. Energy, 151, 420-429.

Hylleberg, S., Engle, R.F., Granger, C.W.J., Yoo, B.S. (1990), Seasonal integration and cointegration. Journal of Econometrics, 44(1-2), 215-238.

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), Grouped functional time series forecasting: An application to age-specific mortality rates. Journal of Computational and Graphical Statistics, 26(2), 330-343.

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.

Liebl, D. (2013), Modeling and forecasting electricity spot prices: A functional data perspective. The Annals of Applied Statistics, 7(3), 1562-1592.

Qu, Z. (2011), A test against spurious long memory. Journal of Business and Economics Statistics, 29(3), 423-438.

R Development Core Team. (2010), R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Ramsay, J.O., Dalzell, C. (1991), Some tools for functional data analysis. Journal of the Royal Statistical Society. Series B (Methodological), 539-572.

Ramsay, J.O. Silverman, B.W. (1997), Functional Data Analysis. Springer Series in Statistics. Springer, New York.

Ramsay, J., Silverman, B.W. (2005) Functional Data Analysis. 2nd edition. New York: Springer-Verlag.

Ramsay, J.O., Silverman, B.W. (2002), Applied Functional Data Analysis, Methods and Case Studies. New York: Springer-Verlag.

Saloux, E., Candanedo, J.A. (2018), Forecasting district heating demand using machine learning algorithms. Energy Procedia, 149, 59-68.

Shang, H. (2013), Functional time series approach for forecasting very short-term electricity demand. Journal of Applied Statistics, 40(1), 152-168.

Ullah, S., Finch, C.F. (2013), Applications of functional data analysis: A systematic review. BMC Medical Research Methodology, 13, 43.

Vilar, J.M., 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.

Yuan, X., Sun, X., Zhao, W., Mi, Z., Wang, B., Wei, Y.M. (2017), Forecasting China's regional energy demand by 2030: A Bayesian approach. Resources, Conservation and Recycling, 127, 85-95.