ELECTRICITY CONSUMPTION PATTERNS IN HOUSEHOLDS

Fermín Mallor[(a)], José Antonio Moler[(b)] and Henar Urmeneta[(c)]

Departamento de Estadística e Investigación Operativa, Universidad Pública de Navarra. Pamplona, Spain

[(a)] mallor@unavarra.es, [(b)]jmoler@unavarra.es, [(c)]henar@unavarra.es,

ABSTRACT

Obtaining patterns for electricity consumption in a particular household is a key point to simulate and to dimension the electricity supply needed in an isolated house. Electricity consumption profile of a user is a function that indicates the electrical consumption in a dwelling over a period of time, usually one day. When this function is considered as a datum and some days are observed, a sample of functions is obtained. Functional Data Analysis (FDA) provides procedures and techniques to analyze this kind of samples. In particular, usual estimators and some procedures of the classical statistical analysis are extended to this context. In this paper we make use of the FDA to analyze the variability in the electricity consumption profiles to obtain consumption patterns useful for simulation of electricity demand in individual households.

Keywords: electricity consumption pattern, functional data analysis, electricity demand simulation.

1. INTRODUCTION

Electrical demand modeling is quite usual in the study of electrical consumption because this analysis is crucial for making decisions about electrical production.

End-user models describe the electrical consumption of a particular household. Depending on relevant characteristics of the consumers, different demand models can be considered. A careful aggregation of the consumer models provides the conventional global electrical demand curve. This procedure is known as bottom-up demand modeling approach. The main difficulty of this procedure is to obtain enough information, in some cases with a high level of detail, from the individual users to design accurate models. Several approaches have been considered in the specialized literature to provide realistic models, see, for instance, Paatero and Lund (2006) or Muratori et al. (2012) and the references therein.

On the other hand, top-down models use global information, say, macro-economic variables joint with global estimates of the energy consumption and structural characteristics of the dwellings, to assign a pattern of the electrical consumption in a particular household. Econometric models have usually been used in this approach. An interesting review about top-down and bottom-up models is presented in Swan and Ugursal (2009) where pros and cons for each methodology are analyzed.

This work is a part of a more ambitious project that aims at obtaining a simulation model of the individual household electrical consumption. This is needed in order to properly dimension the energy supplies that an isolated household may need to cover its necessities. Both methodologies, bottom-up and top-down will be used to reach this goal. On the one hand, big national surveys carried out by statistical agencies are used to establish several profiles of electrical consumers about the total energy in a year. In this work we make an exploratory analysis to identify which general characteristics cluster properly the population with respect to the consumption of energy. On the other hand, data are taken from individual households to establish individual profiles, and observe how to deal with these information in order to determine which aspects explain better the consumption of electrical energy in a particular household. A datum now is the vector of dimension 1440 which contents daily electrical consumption minute by minute of a user and, from the mathematical point of view, these are the values that take a function in points equally spaced on the interval [0,1440]. This function is the daily individual profile load curve of a user. For each user, we observe N days, so we have a sample of N individual profile load curves. In order to handle this kind of data, we appeal to a specific statistical technique which is Functional Data Analysis (FDA) and becomes the natural way to handle this kind of data, see Ramsay and Silverman (2005).

The classical techniques to deal with electricity prices and loads are time series models, see, for instance, an overview in Weron (2006). FDA is seldom applied in this context, however, some papers have appeared in the recent literature that use it with different goals to the one considered here. In the seminal paper by Hyndman and Ullah (2007) FDA is introduced to forecast time series data, even though they illustrate the methodology with demographic data, the extension to other contexts is immediate. In Andersson and Lillestøl (2010) two FDA techniques are presented, functional analysis of variance (FANOVA) and a functional autoregressive model (FAR), in order to make, respectively, data exploration and forecasting electricity consumption. The former helps to study seasonalities and the latter focuses on the time-series nature of the consumption. They consider that FDA is a promising way to search for the data-generating mechanisms in the electricity market. In Goia et al. (2010) FDA is also used for short-term peak load forecasting. They have hourly observations of the aggregate consumption in a district-heating during 198 days in four different years. A basic model is to consider a functional regression model where the response is the daily peak of heating and a functional regressor which is the load curve of the previous day. The forecasts are improved when load curve are clustered with a FDA technique and the basic regression model is applied in each cluster that concentrates load curves that exhibit similar consumption. Recently, in Liebl (2013), FDA is used to model and forecast electricity spot prices but, as stated there, the techniques to model and forecast spot prices are more complex than those needed for modeling and forecasting electricity demand.

In this paper, as stated before, we deal with general surveys on the consumption habits of the population, as the HBS in Spain, see INE (2010), which provides household profiles with respect to the electric consumption. Then, individual load curves for representative dwellings of each profile are sampled where observational points of the daily load curve are taken every minute with a smart meter. In the previous references, aggregated data are used in the applications and this entails smoother curves than individual ones. The variability of individual load curves require some treatment of the sample information in order to apply FDA techniques. In Chaouch (2014) the goal is short-term forecasting of the household-level intra-day electricity load curve so the setting is quite similar to this work but nor the goal neither the statistical techniques used are equal. Guardiola et al. (2014) is an important methodological reference for this paper, although the setting is quite different to ours.

A common conclusion in the previous references is that FDA allows making an integral treatment of the daily consumption and avoids the separate treatment that classical techniques require for the observational points in the same day. FDA also allows us to establish daily patterns depending on the consumer and some environmental characteristics.

This paper is organized as follows. In section 2 a description of the procedure used to collect and organize the sample information is given. In section 3, statistical methodology applied to sample data is briefly explained. In section 4, some results are outlined for a particular household in order to illustrate the technique applied.

2. COLLECTION OF DATA

As a preliminary step, a study to determine typologies of households depending on their consumption of energy has been carried out. This study considers a huge amount of data provided by the Spanish Statistical Institute in the Household Budget Survey (HBS), see the methodology followed in this Survey in INE (2010) . This survey is made every year on 24000 dwellings randomly selected. Each household provides detailed information of their consumption expenditure during two consecutive weeks in each of the two years of their participation. They also fill a form where the consumption of goods which are periodically paid, as the energy, is reflected, therefore, Spanish Statistical Institute publishes every year the annual consumption of each type of energy in the sampled households.This survey considers several classifications to understand the different typologies of households and the classification that shows the best ability to distinguish profiles of total consumptions among households is taken as a benchmark. In fact, a household representative of each class is considered in order to make an individual study. The goal now is not so much to know the total daily consumption, but to be able to simulate consumption curve minute by minute of the day.

For this individual study an energy meter has been installed in several dwellings. The global electricity consumption in the house is saved every minute. So that, not distinction among the appliances used in the household is collected. The information was recorded for one year. The information provided by the device is given in an excel file with two main columns: date and time of the record (DD/MM/YYYY HH/MM) and kwh consumed.

Considering a day as a natural period of time we have a sample of size 365 where each datum is a vector of 1440 components (the electrical consumption of each minute of the day). This sample of functional data is the basis for the study in each household.

The excel file is read by the R software. Date and time information is used to obtain several factor variables: day of the week, month and period of the day.

A new file is built where each day becomes a column (or row) with 1440 observations which represent the kw consumed each minute of the day. The fda package for R fda.usc, see FebreroBande and Oviedo de la Fuente (2012), is used to make an exploratory analysis of these data.

First of all, we should refine the file because some days show extremely low consumption levels. This can be produced because the family leave the house for a long period, for instance, holidays but the most disturbing situation is when some accidental failures of the energy monitor produce large sequences of zero consumption which distort dramatically the analysis. In order to do this refinement, two procedures are followed

a) The routine implemented in fda.usc to detect outliers.
b) A particular routine is implemented in order to detect those days with a high percentage of minutes without consumption that has not been large enough to be considered as an outlier.

This refinement must be careful because we do not have to punish households with efficient appliances that save energy.

3. STATISTICAL METHODS

Data are collected as n observed digitized curves $\{x_i(t_j): j=1, ..., p\}$ with $i=1, ..., n$. The observation points, t_j ,are equidistant. In our case $n=365$ and $p=1440$. This entails a large amount of observations for which classical statistical techniques are not designed to cope with. FDA fills this gap.

The first task in FDA is to convert these data to a function x_i with values $x_i(t)$ computable for any desired argument t. Two approaches are possible: interpolation or smoothing. The latter is better when the target is to clustering data. This is our case because different behaviors of the individual electricity consumption are expected depending on the weekday, the month or other characteristics associated to each day. See, Hitchcock et al (2007).

In this work, the smoothing procedure consists of representing the function $x(t)$ as a linear combination of K known basis functions. This functions are polynomial segments jointed end-to-end at certain argument values called knots. This technique is computationally intensive but, as stated in the previous section, appropriate algorithms have been implemented in the statistical package R to ease the application of the FDA.

An outlier technique implemented in the fda.usc package has been used to eliminate from the sample those days which reflect failures in the energy meter, see more technical details in Febrero-Bande and Oviedo de la Fuente (2012).

There is also a natural extension of the classical summary statistics mean and standard deviation in the FDA framework.

We can also consider with FDA a technique similar to the multivariate technique of principal components, this is called with the acronym PCA. In this case, each functional datum (each day) is written as a linear combination of a basis of splines. The size of the basis is chosen with an optimization function included in the fda.usc package. Then, the role of the components is taken over by the harmonics. Each harmonic is a function which collects some essential characteristic of the electrical usage in a household. Each day can be written as a lineal combination of the harmonics. The coefficients are called scores. The technique is useful when two or three harmonics are enough to capture a high percentage of the variability of the whole data. The harmonics are ordered depending on the percentage of variance that contain. Then, we can represent each day in a plane attending at their scores in the two main harmonics. The goal is to cluster the whole sample of days depending on the harmonic that best represent them. A similar procedure is followed, in another context, in Guardiola et al. (2014) where the interested reader can find technical details.

Finally, smoothing methods, not necessarily based on splines, are considered in order to distinguish in each daily curve a fixed behavior in the individual electrical consumption and the peaks generated by the use of some appliances in the house as dishwasher, washing machine, iron, microwave…. These peaks account for a large part of the data variability. A good understanding of both processes is a basic step for designing a good simulation model.

4. RESULTS

Here we present a short overview of some graphical and numerical results that can be obtained with FDA techniques.

We consider the sample of a user. First an outlier detection technique is applied and n=216 curves (days) are finally considered for the study.

In Figure 1, mean and standard deviation functions are plotted over the profile generated after drawing the 216 curves. We observe that this user presents the most homogeneous behavior from the minute 1200 (20:00 CET) until midnight and wee hours because the mean function takes greater values than the standard deviation. Besides this, some load curves present large peaks, far away from the mean behavior. These peaks of consumption produces the high variability in the data which is so typical in individual load curve, something that gets mitigated when data are aggregated in hours, see Figure 2 f or this user.

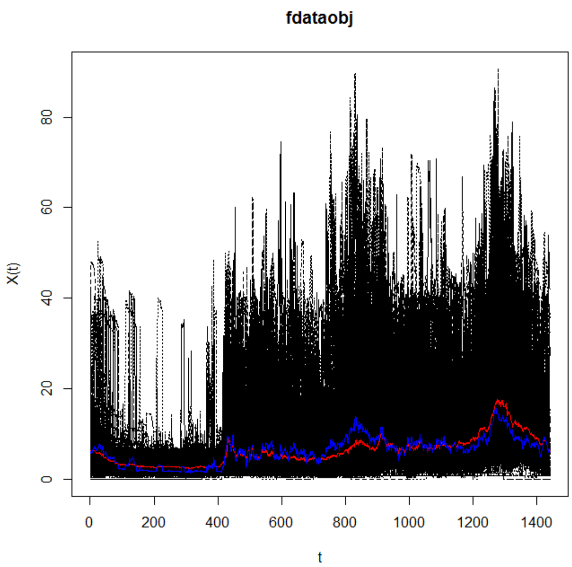

Figure 1. Mean function in red, deviation function in blue. Consumption profile for the 216 curves in black.

A PCA exploratory analysis is applied with the fda.usc package. Unfortunately, the three main harmonics are not able to capture a 30% percent of the variability of the data for any of the households under study. We modify our data file as follows. We consider for each day the electricity consumption ordered in a decreasing way. Now, the graph of each functional data does not represent the amount of consumption at each minute of the day. But it gives the amount of minutes for which the electricity consumption was larger than a specific value, we will call it ordered load curve. This graphic can be obtained for several tracks in the same day. Figure 3 shows in the left panel the ordered load curve when the day is divided in 1 period, and in the right panel, when there is no division in periods.

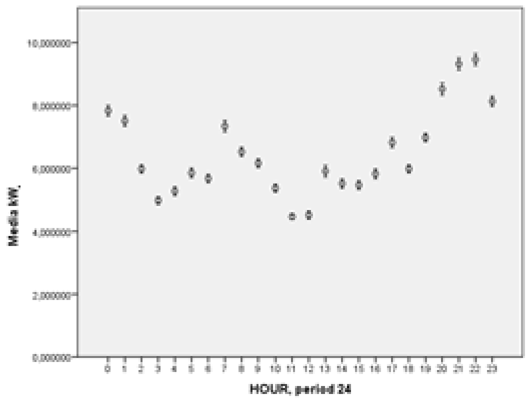

Figure 2. Confidence interval (95%) of the mean consumption per hour

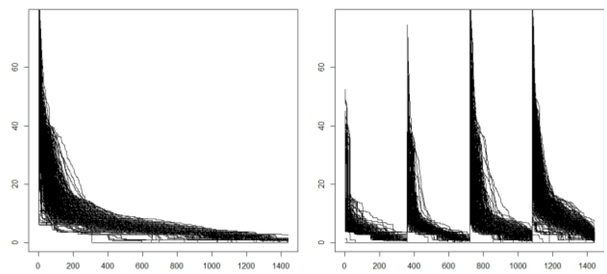

Figure 3. Left panel, Pareto grafic of the electric consumption for the 216 load curves. Righ Panel, Pareto graphic of the consumption for each of the four periods of the day (0,6) (6 , 12), (12, 18) (18, 24)

An analysis of principal components for these new data files provides, for the three first harmonics, an explanation of the total variability that increases, see Table 1, when the ordered load curve only consider one period.

Table 1. Total variability collected by the harmonics.

Data file	Harm 1	Harm2	Harm3	Sum
Load curve	0.1384	0.0817	0.0558	0.2759
Ordered Load curve in four periods	0.4317	0.2352	0.0931	0.76
Ordered Load curve in one period	0.7706	0.091	0.0486	0.9102

The improvement of the power explanation of the harmonics comes from the fact that the ordered load curve in one period does not take into account the exact time when, for instance, one appliance has been used, only takes into account the level of consumption whatever the time of the day. So that, the possible shifts in the habits of a particular household do not affect the ordered load curve, and then a source of variability that is present in the load curve disappears in ordered load curves.

In Figure4 a two dimensional representation of the data is drawn. Days are represented with different colors depending on the weekday. First harmonic distinguishes week-end days from labor days. Similar analysis can be done with other factor variables as month or bank-holidays.

Finally, in figure 5, several smoothing methods are presented with the aim of capturing a basal behavior in the electronic consumption of this family. The peak process becomes, per each day, the values over the smoothed line. Criteria to establish the best smoother are needed.

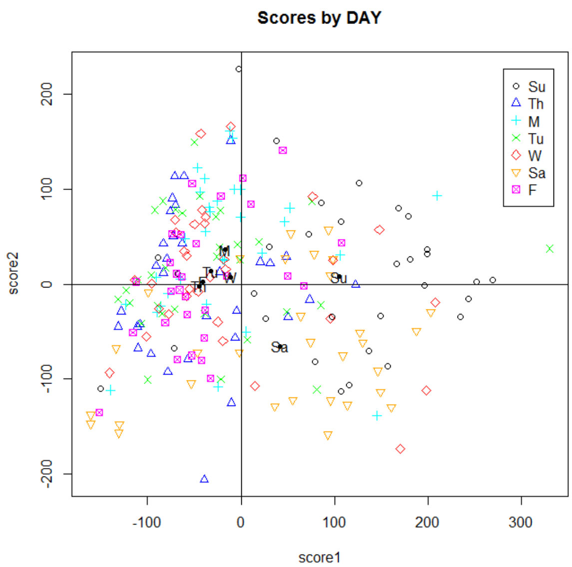

Figure 4. Two axis representation of the functional data distinguished by the weekday.

The HBS classifies households according to some typologies widely accepted by the statistics agencies, see pg. 39 in INE (2010). The main variables that determine a particular typology is the number of people in the dwelling and their age. Attending to the typology D.1.3. , Figure 2 corresponds to a household with a couple with three or more children dependent. On the other hand, Figure 6 shows the sane graphic for a household where the inhabitants are a couple where both are older than 65. Their profiles are quite different because their habits are different.

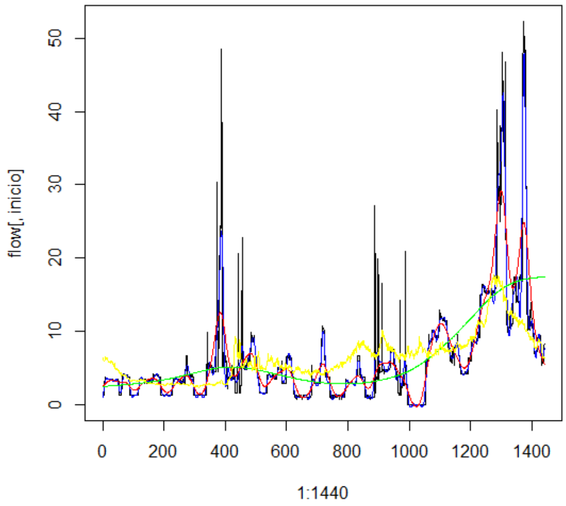

Figure 5. Smoother of a consumption electrical curve for a particular day (black), mean function (yellow), smothers for different parameter for the smoothing algorithm (blue, red, green)

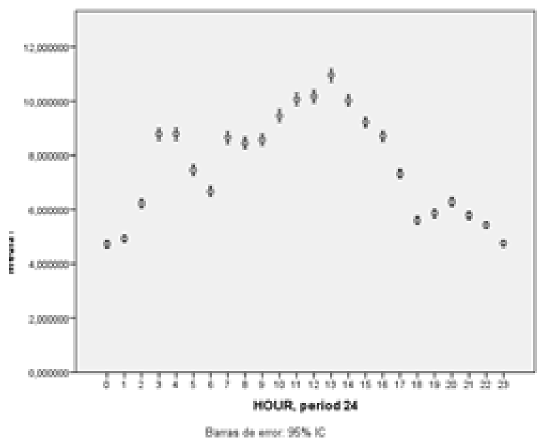

Figure 6. Confidence interval (95%) of the mean consumption per hour in a household of a couple where both are older than 65.

The HBS allows the study, with explanatory goals, of the behavior of the population with respect to the consumption of several sources of energy and the main variables that influence this behavior are the location, the existence of heater and the typology of the household.

REFERENCES

Andersson, Jand Lillestøl, J., 2010. Modeling and forecasting electricity consumption by functional data analysis. *The Journal of Energy Markets*, 3 (1), 3–15.

Chaouch, M. 2014. Clustering-based improvement of nonparametric functional time series forecasting: application to intra-day household-level load curves. *IEEE Transactions on smart grid*, 5 (1), 411-419.

Febrero-Bande, M. and Oviedo de la Fuente, M.., 2012. Statistical Computing in Functional Data Analysis: The R Package fda.usc. *Journal of Statistical Software*, 51 (4), 1–27.

Goia, A., May, C. and Fusai, Gianluca, 2010. Functional clustering and linear regression for peak load forecasting. *International Journal of forecasting*, 26, 700-711.

Guardiola, I.G., Leon, T. and Mallor, F. 2014. A functional approach to monitor and recognize patterns of daily traffic profiles. *Transportation Research Part B, 65, 119-136.*

Hitchcock, D.B., Booth, J.G. and Casella, G., 2007. The effect of pre-smoothing functional data on cluster analysis. *Journal of Statistical Computation and Simulation*, 77, 1043-1055.

Hyndman, R. J. and Ullah, M. S. 2007. Robust forecasting. *Computational Statistics & Data Analysis*, 51, 4942-4956.

INE, 2010, Household Budget Survey, Methodology, at www.ine.es/en/metodologia/t25/t2530p458_en.pdf

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

Muratori, M.; Roberts M. C.; Sioshansi, R.; Marano, V. and Rizzoni, G. (2012)A highly resolved modeling technique to simulate residential power demand. *Applied Energy* .

Paatero, J. V and Lund, P. D. 2006, *A model for generating household electricity load profiles*. *International Journal of Energy Research*;30: 273-290

Ramsay, J.O. and Silverman, B.W., 2005. *Functional Data Analysi, 2nd Edition*. New York: Springer.

Swan L.G. and Ugursal V.I. 2009. Modeling of end-use energy consumption in the residential sector: a review of modeling techniques. *Renewable and sustainable energy reviews*, 13, 1819-1835.

Weron, R., 2006. *Modelling and forecasting electricity loads and prices: a statistical approach.* Wiley.

AUTHORS BIOGRAPHY

Fermín Mallor is Professor in the Public University of Navarra. His research interests are operations research, simulation and reliability.

José A. Moler is Associate Professor in the Public University of Navarra. His research interests are stochastic processes, urn models, experimental designs, and clinical trials.

Henar Urmeneta is Associate Professor in the Public University of Navarra. Her research interests are stochastic processes, probability theory and experimental designs.