An approach to discover the potential for demand response in the domestic sector

Joana Abreu*, Francisco Câmara Pereira[Y], Jorge Vasconcelos[+] and Paulo Ferrão[+]
* MIT Portugal Program Technical University of Lisbon, Portugal Visiting Student MIT, jmabreu@mit.edu;
[Y]MIT Portugal Program at University of Coimbra, Portugal, camara@dei.uc.pt
MIT Portugal Program, IST[+]

Abstract— *Deregulation of the electricity market and electronic metering led to the emergence of retail activity. In particular, the residential sector is the most challenging for demand side management, because the benefits are not immediate for consumers. However, the benefits of the potential increase in efficiency of the residential sector are widely recognized.*
Fears of the risks for the consumers resulting from tariff deregulation, the lack of widespread infrastructure of individual measurement, but also the intrinsic characteristics of the residential sector explain why the majority of consumers are not exposed to real time pricing or other dynamic economic or technologic schemes of demand side management. The work here presented aims to gather information about how the consumers use electricity. Based upon a small scale quasi-experiment, this paper looks to understand whether it is possible to single–out patterns of consumption, identifiable by the electricity consumption signature consistent across the sample. Fourier analysis puts in evidence common periodicities in the data, while factor decomposition aggregates groups of variables: hours of the day. The information extracted from the analysis will be confronted by consumers for an assessment on how far the interpretation is from the actual behavioral trends.

1. Introduction

Electronic meters allied to home area networks present the technical ground for a paradigm shift in demand response in the residential sector. Highly diverse and fragmented [1], the residential sector represents a challenge for demand response, as efficiency gains resulting from a decrease in consumption at the individual level have an impact only when aggregated.

The capacity to aggregate demand and negotiate the residential capacity directly, in wholesale markets [2], can provide an option for new generation and even compete with peak generators.

Electronic consumption sampled in real or close to real time, is gradually becoming available to the academic community, despite important challenges of data security, ownership and general protection of the citizen's privacy.

It is therefore possible to gather knowledge about the patterns of consumption of the residential sector and other information that may inform the more efficient and better accepted demand response programs. The importance of increasing the knowledge about residential research was acknowledged by FERC [3], who recognized the need to provide market operators with accurate assessments and forecasts of demand response to "support just and reasonable rates for the delivery of demand response in wholesale markets, and to accurately measure and verify demand resources that participate in capacity markets."

The process usually led by electricity retailers with consumers aiming to change the shape of demand in periods when energy-supply systems are constrained or when electricity costs are high, is known as demand response. Demand response can result in significant benefits for power systems [4], but can also bring about considerate costs [5] especially if a change in communication infrastructure is in place.

Uncertainty about the recovery of costs and the efficacy of demand response schemes will diminish with the increase of knowledge about specific characteristics of local demand. For a review of a typology proposal of demand response, the reader may refer to [6]. The present paper will focus on the frequency demand response episodes, *i.e.* the frequency at which the aggregator requires the individual household to respond. Such frame opens the ground to justify the identification of the features of residential consumption that can potentially aggregate to cost effective demand response schemes.

To capture the variability of households with respect to high frequency events, a high sampling rate needs to be set in place [7]. However, the sampling rate of the experimental setting supporting this work is the standard 15 min resolution. Although this sample will not show the influence of low period and high frequency events [8], such as turning on and off appliances in the home, the aggregate 15 minute period will nonetheless be sufficient to exhibit variations across consumers, and the presence of periodic equipments.

978-1-4244-6078-6/10/$26.00 ©2010 IEEE

Load modeling has been evolving since the 70's and a review of the methodologies to model the load curve has been provided by [8]. The author basically distinguishes load modeling according to whether the research focuses on long run or short term features of electricity demand. Original studies model the load based on long-term explanatory features such as socio-economic, demographical and residential [9].

The methods look to fit an ARMA time series model while the coefficients of the model are regressed according to the explanatory variables. The main argument against this approach is that residential demand is dynamic, changes with weather conditions [8], and pricing schemes [7]. Recent studies look to integrate this variability though methods that look to mimic actual behaviors through cubic spline curves and discrete Fourier transform.

The analysis presented in the paper applies methodologies traditionally used in econometric studies and in fields of behavioral modeling to highlight patterns of behavior and activity. They are useful to extract knowledge from residential demand and to understand whether the information can discriminate a potential for demand response.

The application of parametric analysis to time series data, where consecutive variables are not independent has been tried before with relevant results [10].

2. METHODS

Data sources and network infrastructure

The work presented in this paper is the consequence of an interdisciplinary study about residential electricity consumption, which includes 15 households measured for 270 days, via an electronic meter that transmits electricity consumption *via* the power line to a communication module, connected by Ethernet to a modem. Every fifteen minutes, the second by second data is aggregated and sent to a central server.

Errors

Data cleaning and preparation involved identifying errors and omissions in the data, which is hardly a novelty in smart metering deployments. With an experience of deploying over 5 million equipments, [11] elects four categories of problems in smart metering deployment: transmission failures (7,8%), networking problems (1%), data measured but not stored (2%) and incorrect installation (4,9%). A fifth category includes error of measurement resulting in incorrectly measured spikes, but the percentage of those errors is close to 0. The experiment underway revealed errors as well. In fact, the nature of the communication process that relies on the technical infrastructure of the resident to transmit the data to the central server influences the errors observed, in particular when the communication module is disconnected from the modem or the modem is turned off over the period of storage.

Predominantly, transmission failure is the most frequent error (66%) but those errors have a less impact in the quality of the dataset, and account only for 1% to 3% omissions. Networking problems are less frequent (20%), yet represent 13% omissions in the dataset. Less predominant problems are what can be described as equipment malfunction or permanent disconnection from the resident. The abandonment rate was 13%.

Table 1: Errors of measurement

% of error in the dataset	**Description of the error**	**Predominance (observed frequency)**
Between 1% and 3%	Transmission failure	66%
13%	Networking problem	20%
Between18% 25%	Equipment malfunction or permanent disconnection	13%

Data preparation

The analysis discussed in the present work looks at the datasets as a whole, per individual household, irrespectively of seasonal variations, because the exercise intended to find repetitive behavior in the data despite the influence of the weather.

The datasets were verified for evidence of normality and autocorrelation. Scatterness and high variance around the mean was corrected using a logarithmic transformation. As expected there was a strong evidence of autocorrelation between consecutive hours: figure 1 shows the evidence of errors in the data. Omissions and zero values were corrected using least squares interpolation (4 neighbors).

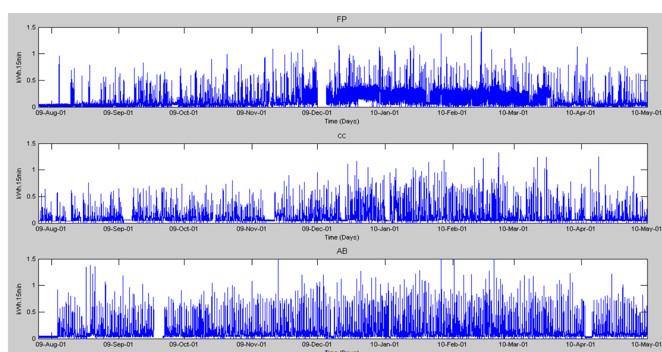

Figure 1: Evidence of errors in the datasets

Frequency analysis

Spectral analysis decomposes the time series into a spectrum of cycles of different lengths, or frequencies. This analysis developed in the frequency domain rather than time, shows how much the signal lies over a range of frequencies. The mathematical function used was the Fourier transform, and the Welch algorithm, whereas the sampling frequency is an hour (96 values).

Eigendecomposition

The principal components analysis (PCA) method transforms a set of p variables (y_1, y_2, …, y_p) into linear combinations of other k variables (z_1, z_2, …, z_k) called the principal components, which will explain a large amount of variance with a smaller number of variables (when k<p). More than just a reduction of dimensionality, it becomes possible to identify trends in the signal. The number *k* determines how much variance is to be explained. The linear combination of the *k* principal components allows the re-representation of vector *Y*:

$$Y = a_1Z_1+a_2Z_2 + \dots +a_kZ_k+\varepsilon \qquad (1)$$

where a_k represent the linear combination coefficents, Z_k the principal components, or eigenvectors, and ε the unexplained variance. Typically, the number *k* is chosen to make ε neglectable.
The analysis was conducted on the correlation matrix and we used the Kaiser rule and the observation of the scree plot [11] to determine the number *k* of components to extract, whereas the original data was transformed using a lognormal transformation, as discussed previously. The advantage of this analysis is to condense the information that composes the signal of electricity consumption of a household (figure 2) in a series of vectors that explain the majority of the variance of the dataset. The eigenvalues demonstrate the percentage of explained variance of the factors, while the eigenvectors, scaled up to the original power consumption (kW), will represent the main vectors that account for the majority of the variability in the household, around 70%.

The subsequent step looks to interpret the variables (hours) that better explain the variance per factor. Therefore, the rotation of the factors was conducted using a Varimax (orthogonal) transformation. This procedure highlights the time periods that better replicate the behaviors.

3. ANALYSIS

For the same households, the spectral analysis of the periodicity in the data was done to different time windows: 60 and 270 days. The higher frequency period was the daily period (23,2 h), followed by the semi-daily period (figure 3). The analysis to shorter length periods puts in evidence events with a periodicity of 2 or 3 hours (figure 4) which may be credited to the presence of cyclic consumer appliances such as refrigerators.

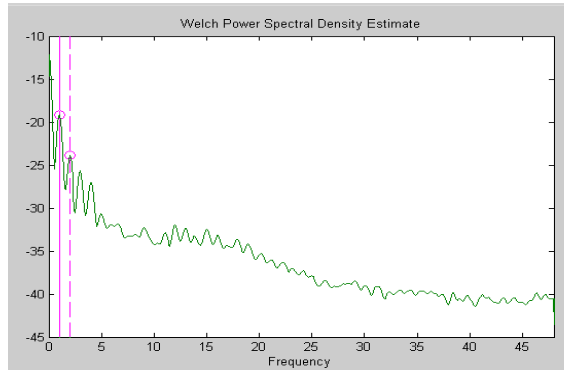

Figure 2: 24 hour and 12 hour periods

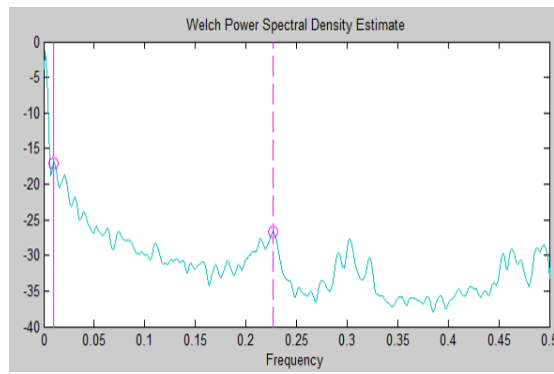

Figure 3: Higher periodicity events such as those provoked by motors of refrigerators are placed in evidence by the analysis.

Having the 24 h period as the validated basis of analysis, we decided to apply the eigendecomposition (PCA) methodology as described in section 2 to vectors of 24 variables, each one corresponding to the aggregated consumptions of one hour. The results systematically indicate that 4 factors are sufficient to explain 70% of the variability of the existing datasets.

Each factor extracted is bounded by the Kaiser rule, which says that the *useful* vectors are the ones with eigenvalue higher than 1, meaning that they fully explain at least one original variable. The following (4 and 5) figures represent two common cases found in the analysis: a case where the first vector explains more than 50% of the variability of the dataset and the case where there the difference between the vectors is not well demarked.

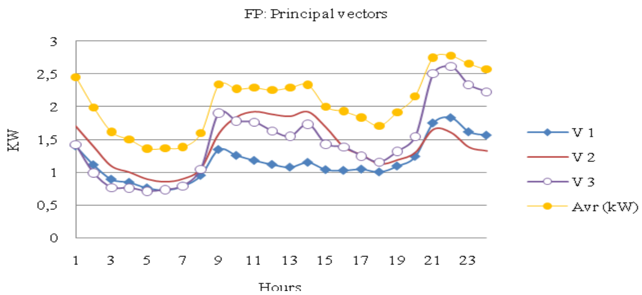

Figure 4: An example of a well demarked contribution of the first vector (V1) for family FP

Table 2: Representativeness of the eigenvalues for family FP

	Eigenvalue	% Total variance	Cumulative %
1	15,7	65,4	65,4
2	1,7	7,2	72,5
3	1,2	5,0	77,6

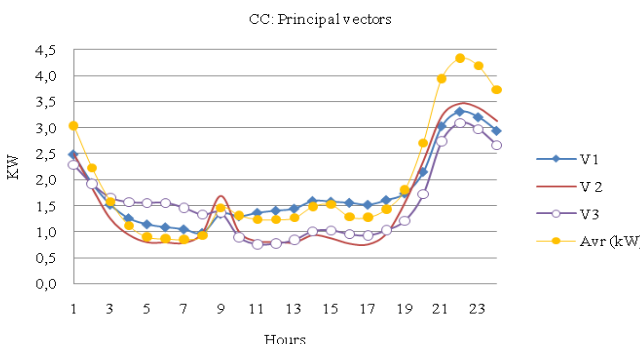

Figure 5: An example of a household where vectors are not well demarked from each other (family CC).

Table 3: Representativeness of the eigenvectors for family CC

	Eigenvalue	% Total variance	Cumulative %
1	6,3	26,2	26,2
2	3,8	15,8	42,0
3	3,1	12,9	54,8
4	1,8	7,6	62,5

The rotation of the axis using Varimax transformation puts in evidence the contribution of the variables to each factor. In this sense, table 4 systematizes the group of variables that make each factor extracted from the analysis. Bearing in mind that the most important factor, in particular, the factor that explains the majority of the variability of the sample, is the first, it is clear that the patterns don't replicate across families for the same factor.

Table 4: Groups of variables by principal factor, per family

	Family	F 1	F 2	F 3
Variable	CG	19 to 21	14 to 16	9 to 12
	FC	22 to 8	9 to 12	19 to 22
	IN	4 to 6	15 to 17	20 to 23
	FP	9 to 14	20 to 23	1 to 6
	CC	10 to 17	19 to 23	3 to 6
	AB	13 to 15	19 to 22	3 to 6
	AV	17 to 23		
	PC	10 to 13	18 to 22	
	MG	0 to 8	14 to 18	10 to 12
	RS	0 to 8	10 to 16	19 to 22

However, the analysis seems to demonstrate a consistency in terms of the variables (hours) that in general correlate more with each factor. As a result, the contributing variables were set chronologically in a table, segmented by the activity periods A through E, as can be seen in table 6, appendix 1. It also shows the standardized distribution of the variables for each factor, and the factors aggregated by correspondence of the activity according to an empirical matrix of classification.

Despite the need for further analysis to investigate the correctness of the results, and to corroborate the identified patterns of activity with the information declared by the occupants of the houses, a preliminary labeling of the periods of activity is presented in table 5.

Table 5: Preliminary classification of periods of activity

Periods of activity	Theoretical explanation of activity
A	Night period
B	Early morning: possibly resulting of the influence of bi-hourly tariff
C	Morning activity: either the influence of housekeeping activities or the presence of teens / elders
D	Afternoon activity: either the influence of housekeeping activities or the presence of teens / elders
E	Dinner time activity

The data shows the persistence of dinner / early evening period for every family involved in this part of the analysis.

4.CONCLUSION

Demand in the residential sector is not sufficiently well understood. As discussed before, the adaptability of demand response programs and smart metering deployments to the general population is starting to be challenged by consumers. Therefore, the success of demand response programs lies on an effective knowledge about the characteristics of the consumers and respective patterns of behavior. On the other hand, the cost effectiveness of the programs will depend on an assessment of the potential for demand response in particular to load shedding, shifting and curtailment underlying a specific population that inhabits the region of interest. The analysis presented in this paper shows consistency in 24 h and 12 hour periods, and that it is possible to identify constant periods of consumption which may be assigned to activity-based models of behavior.

Additionally, the fact that large consumer appliances are visible in the analysis may confirm a potential for improvements in energy efficiency. As probable explanations, the results of the analysis will inform the interviews to the residents since they suggest specific behaviors. Further research will look to the evidences shown by shorter periods (12 h) and will compare with the National loads. Additionally, the authors will look to distinguish and quantify high versus low frequency consumption.

Despite the exercise initiated in this paper, a great weakness is the number of families involved in the analysis. Far extrapolating the results to a larger population, the present research exploits methodologies to analyze residential electricity consumption.

ACKNOWLEDGEMENTS

Special acknowledgement to Prof. David Marks for the support throughout the research, and Fundação para a Ciência e Tecnologia (FCT).

REFERENCES

[1] Farrell. and Remes, J.. "How the world should invest in energy efficiency". *Economic Studies, The McKinsey Quarterly*, 11p. 2008.

[2] Carlos Vasquez, Carlos Batlle, Sara Lumbreras and Ignacio Arriaga, "Electricity retail regulation in a context of vertical integration: the debate on regulated tariffs", *IIT Working Paper* , IIT-06-028a, 19p., 2006.

[3] Federal Energy and Regulatory Commission, "An assessment of demand response and smart metering". *Staff Report*, 2009.

[4] Sadeghi Keyno, Ghaderi, Azade and Razmi, "Forecasting electricity consumption by clustering data in order to decrease the periodic variable's effects and by simplifying the pattern", *Energy Conversion and Management*"***50**, 829-836,* 2009.

[5] Conchado and Linares, "Gestión activa de la demanda eléctrica doméstica: beneficios y costes". IIT-Universidad Pontifica Comillas, 22pp, 2009.

[6] Jorge Vasconcelos, J. (2008) "Survey of regulatory and technological developments concerning smart metering in the European Union electricity market". *RSCAS Policy Papers* , Florence School of Regulation, European University Institute, 2008.

[7] Wright and Firth "The nature of domestic electricity-loads and effects of time averaging on statistics and on-site generation calculations", *Applied Energy* ***84**, 389–403*, 2007

[8] Matteo Manera and Angelo Marzullo, "Modeling the Load Curve of Aggregate Electricity Consumption Using Principal Components Analysis", *Environmental Modeling & Software,* ***20**:11,1389-1400,* 2005.

[9] Wide, Lundh, Vassileva, Dahlquist, Ellega,. "Constructing load profiles for household electricity and hot water from time-use data—Modeling approach and validation", *Energy and Buildings* ***41**, 753–768,* 2009

[10] Francesco Calabrese,.Jonathan Reades and Carlo Ratti. "Eigenplaces: Segmenting Space through Digital Signatures", Pervasive Computing: ***10**, 76-84*, 2010

[11] Shmueli, Patel and Bruce, "Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner", *Wiley*, *298pp*, 2007.

Table 5: An example of behavioral activity classification and its correspondence with households

Family Code	Activity Period A	Activity Period B	Activity Period C	Activity Period D	Activity Period E
CG		5 to 7	9 to 12 and 0 to 1	14 to 16	19 to 21
FC			9 to 12		19 to 22
IN		4 to 6		15 to 17	20 to 23
FP	1 to 6		9 to 14		20 to 23
CC		3 to 6	10 to 17		19 to 23
AB		3 to 6		13 to 15	19 to 22
AV					17 to 23
PC			10 to 13		18 to 22
MG	0 to 8		10 to 12	14 to 18	20 to 21
RS	0 to 8		10 to 16		19 to 22
Contribution of the variables to the factors that constitute each activity period (standardized)