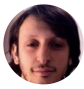
Using pattern recognition to identify habitual behavior in residential electricity consumption

Joana Margarida Abreu
Fraunhofer USA Center for Sustainable Energ…



Francisco C. Pereira
Technical University of Denmark

Energy and Buildings xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

Using pattern recognition to identify habitual behavior in residential electricity consumption

Joana M. Abreu[a,*], Francisco Câmara Pereira[b], Paulo Ferrão[c]

[a] MIT Portugal Program, Centre for Innovation, Technology and Policy Research – IN+, Instituto Superior Técnico, Technical University of Lisbon, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal
[b] Department of Informatics Engineering, Faculty of Sciences and Technology, Coimbra University, Singapore-MIT Alliance for Research and Technology, S16-05-08, 3 Science Drive 2, Singapore 117543, Singapore
[c] Instituto Superior Técnico, Technical University of Lisbon, Av. Rovisco Pais, 1, 1049-001 Lisbon, Portugal

ARTICLE INFO

Article history:
Received 20 December 2011
Received in revised form 18 February 2012
Accepted 28 February 2012

Keywords:
Electricity patterns
Household
Baselines
Routines
Pattern recognition

ABSTRACT

Recognizing habitual behavior and providing feedback in context are key to empower individuals to take control over residential electricity consumption. Yet, it is a challenge to change habitual behavior, embedded in everyday routines. This paper intends to discover whether habitual behavior can be identified by pattern recognition techniques. The data source is an experiment similar to a utility led advanced metering infrastructure implementation.

The analysis discovers: (1) persistent daily routines and (2) patterns of consumption or baselines typical of specific weather and daily conditions. Approximately 80% of household electricity use can be explained within these two patterns, with several applicable "profiles" for this population, including: unoccupied baseline, hot working days, temperate working days, cold working days, and cold weekend days.

The proposed methodology demonstrates that it is possible to use pattern recognition methodologies to recognize habitual electricity consumption behavior given the intrinsic characteristics of the family. This approach could be useful to improve small scale forecast, and as a mechanism to enable the provision of tailor-made information to the families.

1. Introduction and motivation

Energy use by individuals in the privacy of their homes is influenced by very diverse factors. van Raaij and Verhallen [1] in their research in the 1980s, recognize several factors that drive household electricity consumption behavior, such as energy-related attitudes, personality, socio-demographic factors, building characteristics, energy prices, feedback and general information about energy use. In their analysis of the problem, these researchers suggest that routines can become alternative predictors of electricity consumption [1] because routines or habits may resist the cognitive and financial drive and still prevail over rational alternatives [2].

A consequence of the debate about implications of routines in electricity consumption is whether tailored information delivered in context of household habits can contribute to decrease (or shape) household electricity use (or in abstract, energy use). Bridging the gap between technology/information and routines, the analytical approach developed in this paper will look at a semi-automatic method to provide information that can be used to promote new usage habits. Our approach is to benchmark current real time electricity consumption with historical best practices, established by the family for weather and other loading conditions such as weekends, weekdays and months. These experimental methodologies were developed upon the data collected through an experiment, which for the purposes of this analysis closely resembles a utility led automated meter infrastructure implementation. In our case, the experiment includes a total of 15 homes.

The main findings include a method that positively pinpoints the daily routines for a particular household. A second outcome of the analysis is to enable a greater sophistication in feedback mechanisms, which may begin to give back information that can benchmark real time consumption with "normal" consumption for a particular day-type. Individuals could compare their best pattern of consumption with the worst or the average, depending on the objectives of the feedback program.

The underlying theory supporting the research is change of habits, driven by feedback: there is evidence that personalized action-driven feedback is a drive for behavior change for some individuals [2,3]. Our methodology extracts patterns of activity that depend on the loading conditions (i.e.: external conditions such as temperature and working days or weekends). A possible

* Corresponding author. Tel.: +351 933 419 042.
E-mail address: joanamabreu@ist.utl.pt (J.M. Abreu).

0378-7788/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.enbuild.2012.02.044

application of such information is to provide personalized benchmarking against real time consumption.

The inspiration for this type of analysis is developed by Calabrese [4] and Eagle [5]. In the case of Calabrese [4], they use eigen decomposition and clustering analysis to infer human activity from WiFi hot spot sensors distributed across Massachusetts Institute of Technology (MIT) campus.

An example of inferring activity patterns from indirect sensing mechanisms through datamining comes from the works of Eagle and Pentland [5]. In their "reality mining" project, the team uses mobile cell-id and bluetooth data from 100 students' cell-phones as sources of activity, which is then datamined to develop a model of activity and mobility in campus.

Two recent works by Santin [6] and Yu [7] have made use of similar techniques as the one presented in this paper but with completely different approaches. Santin's study is focused on discovering heating habits through survey analysis, while Yu 's [7] study relies on building data and their characteristics (e.g. building signal envelope and type of appliances) to develop clusters of buildings with the same level of electricity consumption. The novelty of the present work is the focus and scope: it focuses on the household and yet keeps the perspective about how the information gathered can be applied to large populations, with the objective of gathering intelligence and capacity to act by providing tailor-made services: tailor made feedback or segmentation.

2. Discovery of the daily routines

The most significant steps of the methodology are explained in the following paragraphs (for an overview, see Fig. 1). The framework comprehends the collection of both quantitative and qualitative household information. The quantitative dataset comprises electricity readings, which comprehend total electricity consumption in 15-min periods. This dataset requires the installation of electronic meters in the households. The qualitative dataset is based upon the responses and narratives provided by the household occupants, when subject to surveys and interviews done by the researcher.

The first step starts with decomposing the signal into its fundamental frequencies, using fast Fourier transform (represented by the box spectral analysis). This analysis highlights the most intense periods, throughout the dataset, and gives evidence of the unit of analysis. For that reason, household time-series were organized in days or groups of 24 h (Fig. 1: unit of analysis).

After grouping the data into the most common periods, the second step is to identify the periods of recurrent behaviors, which occur every day. This discovery process is the result of the application of principal components analysis and rotation of the most significant eigenvectors (Fig. 1: eigenvector decomposition and varimax rotation). The patterns of recurrent behavior, that correspond to the groups of variables that correlate the most with the significant eigenvectors, were matched with the interviews to interpret the results (Fig. 1: match 1).

The final output creates baselines specific of a combination of loading conditions, by clustering the eigenvector coefficients (Fig. 1: clustering). The load curves generated are then confronted with the information collected from the interviews. In this sense, specific patterns of electricity use are confronted with reports of "going on vacation or weekends off" or with evidence of "we began to heat the house" in a specific week or day. Therefore, relying on the interview data we could verify whether the baselines had a correspondence with the household electricity consumption habits in diverse circumstances, throughout the year (Fig. 1: match 2).

To summarize, the methodology proposed is able to create two different outputs:

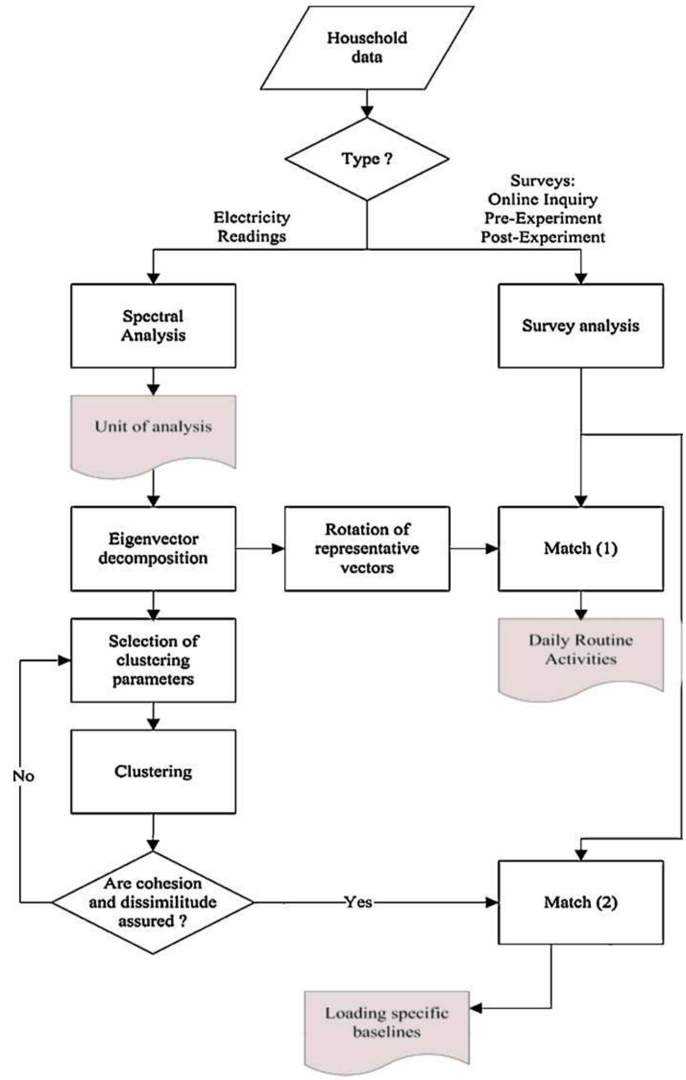


Fig. 1. Diagram explaining the process of discovering the daily routines and baselines specific of loading conditions.

(a) *Daily electricity consumption patterns* – which correspond to every-day routines: the major patterns or recurrent behaviors, common to the majority of the days of the dataset.
(b) *Loading specific baselines* – baselines or routines that are common to groups of days that are characterized by similar loading conditions across the year, which depend on external load conditions. We discover consumption and contextual patterns for specific loading conditions by clustering eigenvector coefficients of daily signals.

2.1. Unit of analysis

Electricity consumption readings were sampled at 15-min intervals. As with earlier work [4], fast Fourier transform was used to decompose the time series into its fundamental frequencies. The frequency approximating the 24-h period is the strongest (or second strongest, see Table 1) and as a result, that period became the unit of analysis. Therefore, a day-vector is composed by 24 readings, which span from 0:00 to 23:00. Secondary frequencies could also be identified, at shorter periods: every 2–3 h. These likely reflect engines that turn on or off periodically (e.g.: refrigerators or air conditioners).

Please cite this article in press as: J.M. Abreu, et al., Using pattern recognition to identify habitual behavior in residential electricity consumption, Energy Buildings (2012), doi:10.1016/j.enbuild.2012.02.044

Table 1
Spectral density estimate- summary results for the first and second periods.

Household code	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
1st period (h)*	24	24	12	24	24	24	24	24	24	24	24	24	12	24	24
2nd period (h)*	12	12		12	12	12	12	12	2	12	12	8			

* Approximate periods.

2.2. Daily patterns of recurrent behavior

Principal component analysis (PCA) is a variable reduction method that builds orthogonal components, i.e., linearly independent with the internal product, or eigenvectors which explain the maximum variance of the original universe. It gave a perspective about the internal structure of the data.

2.2.1. Data interpolation

The electricity consumption for a given hour is highly correlated with the electricity consumption of its immediate precedent and subsequent hour, as seen by the autocorrelation diagnostic in Fig. 2. Considering this, missing data was interpolated using the trend of the past three values to interpolate a missing value.

2.2.2. PCA analysis

For this particular analysis of the daily routines we kept only the eigenvectors with an eigenvalue above 1, as suggested by the Kaiser rule [8] generally giving up to 80% of variance explained. Sensitivity analysis helped make this decision: we were looking for recurrent behaviors that would be maintained throughout the year on a daily basis (and independently of the level of energy used, which differs widely according to the loading conditions).

The structure of the relation between the variables (hourly consumption) and the vectors is not yet evident at this stage of the analysis. Hence, the vectors were rotated to highlight the variables that correlate the most with each eigenvector, and to enhance the relationships between the variables.

A varimax rotation [9] is applied because it maintains the independence of the vectors while highlighting the structure between variables and further reducing the number of vectors needed to describe each day. In practice, each vector focuses on a part of the day that has a significantly persistent behavior, enabling us to isolate routine activity periods.

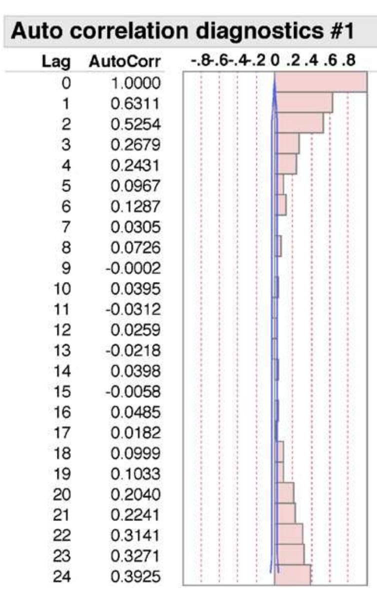


Fig. 2. Autocorrelation coefficients.

This time, each vector gives evidence of the group of variables that contributes the most to its formation. As a result we found persistent patterns that emulate the household's reported routines (i.e., homes #1, #2, #4, #9, #10, #15). For the other houses we have not found such strong evidence. An explanation that applies to households #5, #8, #13 and #14 is that the datasets were too small, becoming sensitive to seasonal routine change (see Table 2).

If the sample is small enough, every component of a day can be different each day, and when the days start to build up, it is expected that these daily routines adopt a common trend. Yet the explanation for households #3, #6 and #7 is not as straightforward. These three households reflect families of three (#6 and #7) or four (#3) where the parents live with one or two children.

2.2.3. Confronting data analysis with actual reports

Next we evaluate whether these recurrent activity periods correctly describe the routines in the lives of the experiment participants. In-depth interviews with target participants facilitated the interpretation of these periods. The interview questions were designed with the objective to prevent the manipulation of the participant responses. In particular, the participants were asked to describe in detail the routine activities of household members before being shown their electricity consumption analysis or having answered any question originated from this analysis. Table 3 gives an example of the degree of conformity between narratives of the individuals and the results computed.

Family #1 shows that despite of its eigenvalue the capacity of the eigenvectors to explain the majority of the behavior of the household was low (i.e., around 50% total). Table 3 also shows the PCA's persistent daily behavior periods and their interpretation given the responses made by the occupants. During the night these periods were so short that they could not be well explained unless we connect the activities in the night period with those of the morning: since the dishwasher is turned on after dinner, it could be expected that its cycle would be prolonged through the night. That, combined with electrical heaters that stay on in the early hours of the evening can explain why the overnight period is so short.

A comparable analysis was applied to the remaining households, and the persistence periods were confronted with the information reported. So far we have discussed the application of the PCA in datasets that consist of the readings collected by smart meters. For a group of households, and using only the significant vectors, we verified that the periods of recurrent behavior identified in the analysis match the information provided by the occupants when asked to describe the routine of a day for every family member. The next step of the analysis looks at these routines across the year.

3. Baselines as an attribute of specific loading conditions

Everything else remaining equal (e.g.: the household keeping more or less same electrical appliances, and the number of occupants for a time period) is it possible to cluster "typical" baselines that depend on the loading-day conditions (e.g.: season, external temperature or day of the week).

We applied the clustering analysis to the eigenvector components representing 80% of the variance in the dataset. The cluster analysis is based on K-means and hierarchical clustering. Essentially, we apply the latter to determine the appropriate "K" value

Table 2
Number of reading-days by household.

Household code	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
Number of days	434	441	377	441	112	432	446	185	359	308	428	523	108	67	408

for *K*-means. For each case we look for the ideal parameters, but followed the general criteria of having meaningful lower and maximum limits:

(a) We chose fifteen as the lower cluster size hoping to combine days, in one cluster, atypical of the rest of the year (due to heat/cold waves or special days). We found that this condition would still keep out odd days or outliers. Decreasing the limit even further would fragment the clusters, and would deter pattern analysis.
(b) We chose one hundred and fifty as the higher cluster size in the hope of clustering together "normal behavior for temperate days". Exceptions were made for small datasets, for which it made sense to separate clusters that would otherwise be kept together: hence the decision was to evaluate small differences given the fact that the duration of their experiment was shorter. The maximum and minimum cluster size, in this case, was eight and thirty, respectively.

The minimum cluster size that initializes *K* for the *K*-means clustering algorithm is six – for similar sized datasets the clustering conditions were not adapted to each dataset. A post-clustering analysis is then performed where each day is classified according to loading conditions. The attributes (or loading conditions) that classify each day are: average daily temperature, weekends and official holidays; these were bundled together and constitute another descriptive characteristic of the data: working days and month of the year. As an example, the clustering results for household #1 are discussed in detail in the next sections. A full discussion about the whole household group is summarized in Section 4.

3.1. Household #1

For household #1 the clustering conditions are Euclidian distance of 0.2 and minimum and maximum cluster sizes of 15 and 150 days, respectively. The process to select these limits was iterative. However, there were also other reasons based in the fact that in one hand we did not want to isolate a cluster so small that would have limited probability of being repeated again. The maximum limit was included almost a third of the days present in the dataset and approximately 5 months seamed large enough to describe the majority of the variability present for specific loading conditions (e.g. temperate days or weekends). The number of clusters was robust because a change in the parameters would not significantly affect results. Six eigenvectors are representative of 80% of variance in the data. The dataset includes 434 days of 24 hourly electricity readings. The result of clustering the eigenvector coefficients generates five clusters (Fig. 3). In a brief overview, cluster 3 comes forward because it represents 35% of the dataset and shows a steady low consumption baseline, cluster 4, groups significantly less (9%) but it gives evidence of high nocturnal consumption occurring mostly over weekends during winter, while cluster 2, represents 11% but differs from the remaining clusters by its higher electricity consumption during afternoons.

To interpret the clusters, each day of the dataset was classified *a priori* according to maximum daily temperature, day of the week (weekend – including official holidays or working days), and month. A graphical representation of a cluster comes along with information about the number of days clustered (cluster size) and with a load curve showing the average value of consumption and corresponding error deviation to the consumption centroid for each hour.

This graph comes alongside with three other pictures that provide the context for each cluster: the distribution of maximum daily temperatures found in the cluster (on the right of the cluster graph), a pie chart showing the proportion of the number of days per month, and a pie chart showing the proportion of weekends against working days. For household #1, a cluster by cluster discussion will be done next.

3.1.1. Cluster 0

As shown in Fig. 4, 151 days compose cluster 0. It reflects a load curve of low to medium electricity consumption. The relative variation of the error about the cluster centroids is small. It is more representative of warmer days (the maximum daily temperature is 23 °C), and in particular, it occurs mostly during summer, autumn and spring. The days included in this cluster are well distributed amongst weekend and weekdays. The cluster seems to represent "electrical consumption" for a warm temperate day, but includes temperatures typical of the summer (maximum of 40 °C).

3.1.2. Cluster 1

As shown in Fig. 5, cluster 1 aggregates 49 days and is particularly significant of electricity consumption in the winter. The evening peak is wider than in cluster 0, and it comprehends more winter days. It stretches overnight and could represent moderate space heating and the functioning of large appliances to make use of the electricity savings period. The temperature distribution is yet skewed to the left (high frequency of cooler temperatures, between 10 °C and 19 °C). The error across the hourly centroid curve is larger in particular in the late evening. It seems to correspond for the load of "temperate days" for moderate outdoor temperature.

3.1.3. Cluster 2

Cluster 2 groups 49 days that are more common in the autumn and early summer (Fig. 6). The most relevant evidence of load curve is its high consumption during afternoons, which could represent days when the household is occupied after lunch, when the laundry and dishwashing machines are left on or even when air conditioning is left on to cool off the house for the when the occupants return home after work. These options were all mentioned as reported behaviors by the occupants of the house. Its relevance in the cooler months of November and December might be representative of

Table 3
Periods of recurrent behavior for household #1 (composed by 2 working adults).

Eigenvalue	Tot. var. (%)	Period	Routine activities as reported during the interview
6.42	26.73	4–6 AM	Persistent period that happens when there is no electrical activity during the night.
3.11	12.97	3–5 PM	Repeatedly, one of the occupant's comes home for lunch and turns on the washing machine. "Bimby" (a self cooking pan) stays on during the afternoon to prepare dinner
2.54	10.58	8–11 PM	Both occupants arrive late at home; dinner, dishwasher.

Please cite this article in press as: J.M. Abreu, et al., Using pattern recognition to identify habitual behavior in residential electricity consumption, Energy Buildings (2012), doi:10.1016/j.enbuild.2012.02.044

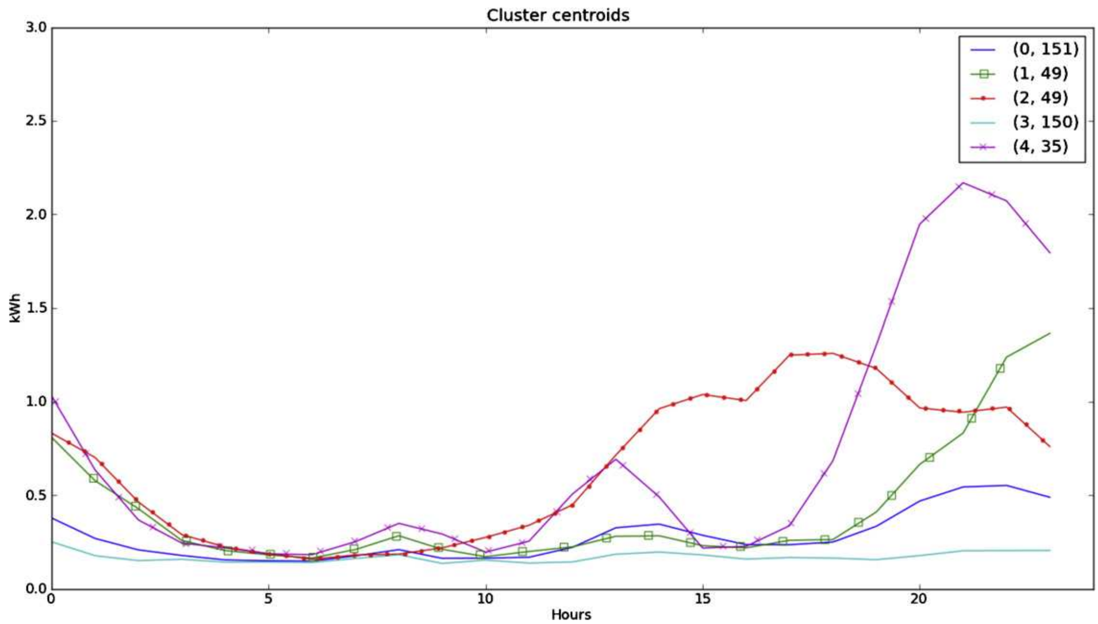


Fig. 3. Cluster centroids, household #1.

similar behaviors, only for heating purposes. Yet there is a larger incidence base for the temperature distribution.

3.1.4. Cluster 3

As shown in Fig. 7, cluster 3 is large cluster that includes 150 days, or about 34.6% of the total number of days. It embodies the baseline of the house, in particular, the load curve that occurs when the house is unoccupied. The appliances that make this baseline are cooling appliances (refrigerators) and other devices left on stand by or off mode. The temperature distribution chart is robust, mostly clustered under the normal distribution line, but the base temperature chart is large, which means the profile occurs almost independently of the temperature, yet, it is more frequent in warmer months. The distribution across months is wide, but more the cluster is more frequent in the spring, early autumn and especially summer. The household is much less likely to run the home at a baseline level in the winter months.

3.1.5. Cluster 4

The number of days grouped by this cluster is 35 (Fig. 8). It is characterized by high electricity consumption in the evening. The temperature distribution is skewed to the right, instead of showing a robust middle term variation, as would the distribution that characterizes cluster 3. The irregular distribution of the temperature chart suggests larger variation of consumption, which is reflected by larger error about the cluster centroids. We see increased consumption in mid-day (between 13:00 and 14:00) and evening (between 19:00 and 02:00). The month range is between November and February, with weekend and holiday day-type days proportionally high in comparison to weekdays. The load curve is

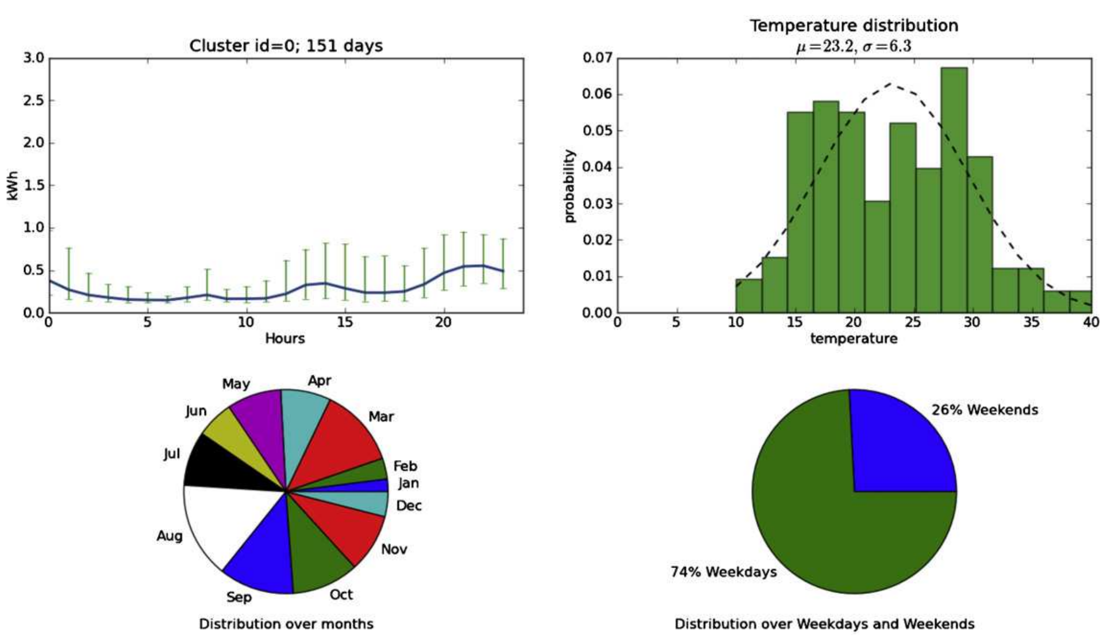


Fig. 4. Loading conditions and load form for cluster 0.

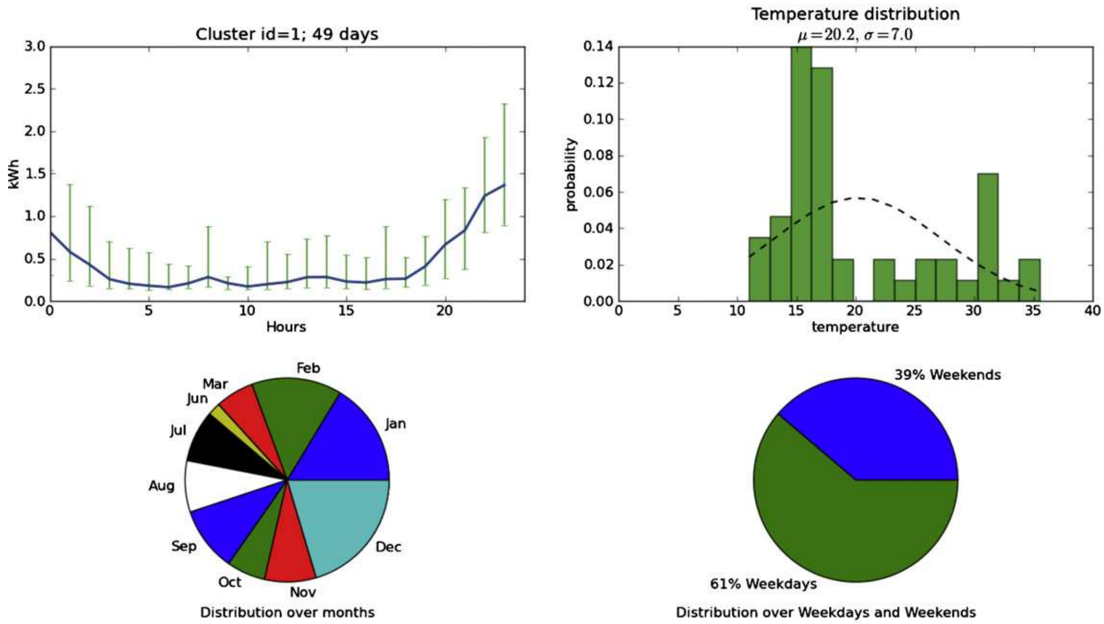


Fig. 5. Loading conditions and load form for cluster 1.

characteristic of this family, for cold weekend days (average temperature 14 °C).

4. Summary of the results

Depending on the dataset size, the households were classified in two: large sized group (up to 14 months of data) and small sized group (up to 3 months of data). For each group, the clustering conditions were different. Each cluster was therefore classified in a matrix that tried to condense the variability of clusters. There are two main groups of households in the analysis: large dataset and small datasets. Both used specific clustering conditions, and are showing distinct effects: the large dataset highlights yearly trends, while the clustering done to the smaller datasets put in evidence shorter term distinctions and it is possible almost to identify a specific loading curve per month. These effects are shown in Tables 4 and 5. The baseline stands out as one of the most common profiles. It is more frequent to distinguish between winter weekends and winter and working days, while that distinction is almost not possible to make in the summer. This is possible due to specific characteristics of the load in Portugal, because HVAC residential penetration is so low. It will be interesting to test these methodologies with other climatic regions where HVAC penetration is higher and heating is still electrical. For the context of this experiment, it is also fairly common to find "temperate cold", "medium temperature weekends" and "temperate warm" profiles.

It is also possible to verify variations in the patterns happening in temperate days (slightly colder or warmer, and with the possibility of occurring throughout the year).

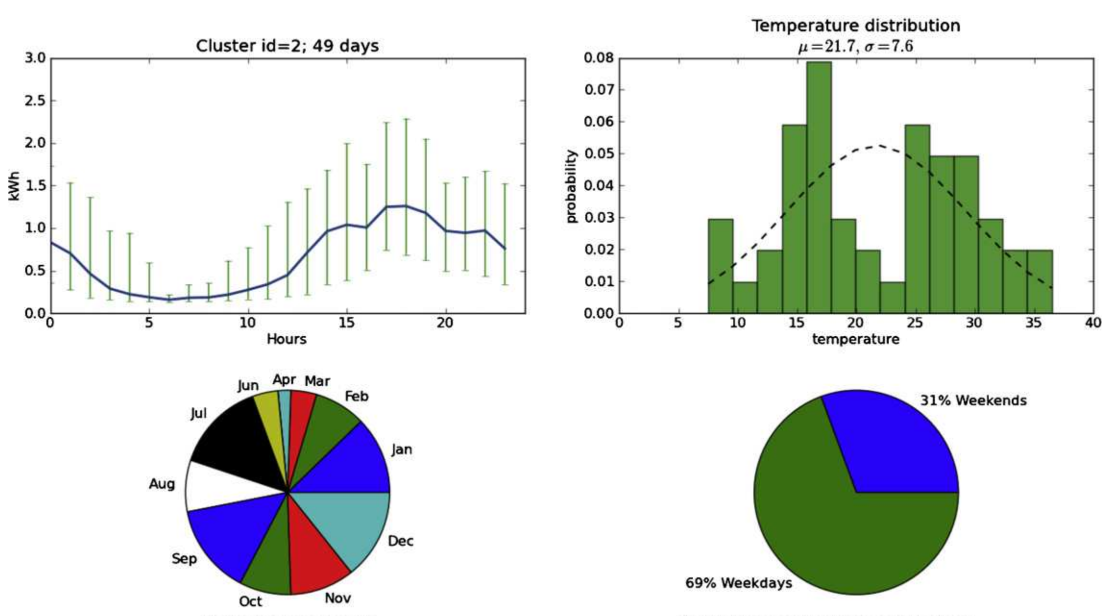


Fig. 6. Loading conditions and load form for cluster 2.

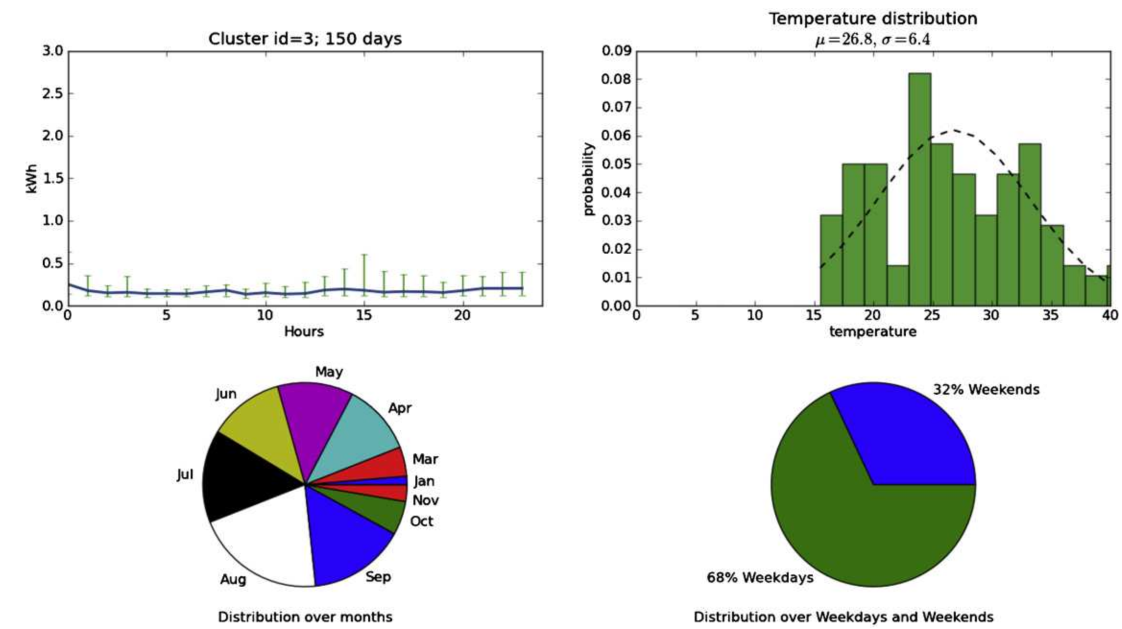


Fig. 7. Loading conditions and load form for cluster 3.

Table 5 summarizes the most common profiles that result of applying the methodology to smaller datasets. In this case, the technique can help distinguish load profiles typical of a month, and at some extent, even differences between working days and weekends in each month.

5. Limitations

Autocorrelation can become a limitation in time series analysis. Using eigenvector coefficients each day is represented by a combination of time series signals. As a consequence, we are no longer comparing similar time attributes, but instead we compare whole signals. This fact makes the analysis more robust.

Another limitation of the process is using average daily outdoor temperature as an attribute. The adaptive comfort literature [10] proves that degree days or the variation of the temperature of the day according to a value previously established according to the season and climate and geography of the place should be tested in place of temperature because humans adapt and do not make the same decisions (of heating and cooling the house in different seasons for the same temperature). It is also possible that using heating degree-days and cooling degree-days as classifiers will improve interpretation (decreasing the number of climate related bins to interpret).

A limitation of the study is the lack of interpretative explanation about which habits and appliances should be responsible to

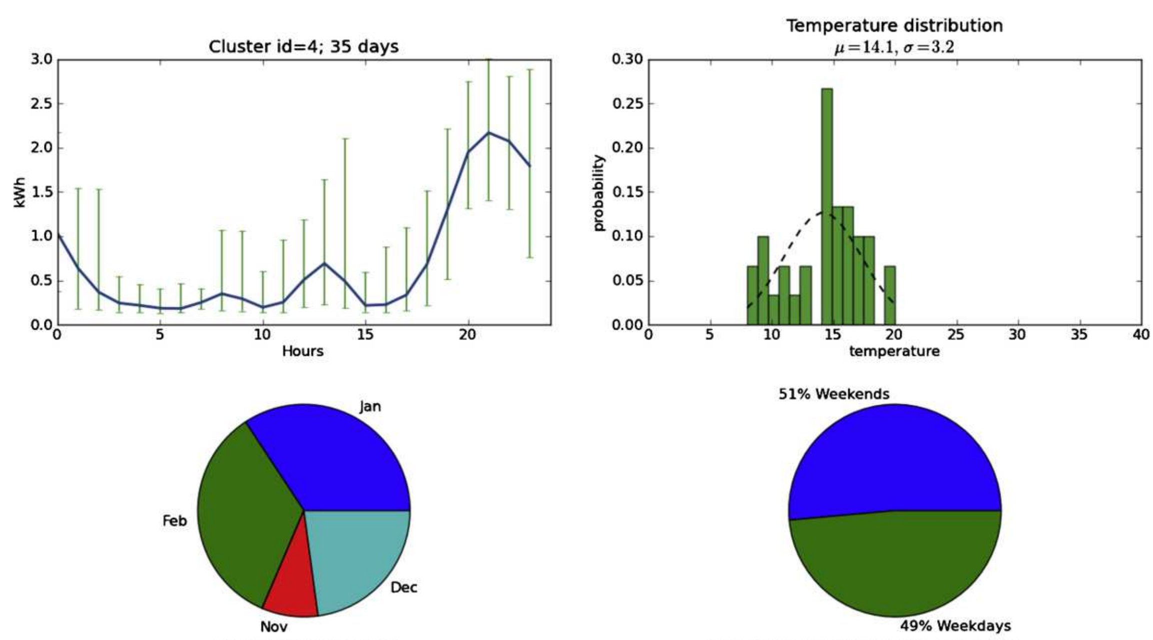


Fig. 8. Loading conditions and load form for cluster 4.

Please cite this article in press as: J.M. Abreu, et al., Using pattern recognition to identify habitual behavior in residential electricity consumption, Energy Buildings (2012), doi:10.1016/j.enbuild.2012.02.044

Table 4
Distribution of clusters across consumption profiles: large datasets.

Hsh.	Winter			Temperate weather									Summer			Baseline
				Colder temp.			Medium			Warmer temp.						
	Wkd	Wkr	EqP	Wkd	Wkr	EqP	Wkd	Wkr	EqP	Wkd	Wkr	EqP	Wkd	Wkr	EqP	
#1	x					x					x	x				x
#2		x	x				x					x			x	x
#3	x						x		x							x
#4	x					x			x			x			x	
#6	x	x							x	x	x					x
#7	x				x		x			x					x	x
#9	x	x	x			x			x							x
#10			x						x							
#11	x	x		x								x	x			x
#12	x	x					x	x		x						x
#15	x	x				x			x							x
Freq.	9	6	3	1	1	4	4	1	6	3	2	4	1	0	3	9
Rank	1st	2nd	4th	6th	6th	3rd	3rd	6th	2nd	4th	5th	3rd	6th	7th	4th	1st

Hsh: household code; Wkd: increased proportion of weekend and holidays; Wrk: increased proportion of working days; EqP: approximately equivalent proportion between weekend and weekdays (eq. proportion is 71%:29%).

Table 5
Distribution of clusters across consumption profiles: small datasets.

Hsh	August/September			September/October			October			November			December			January			Baseline
	Wkd	Wkr	EqP	Wkd	Wkr	EqP	Wkd	Wkr	EqP	Wkd	Wkr	EqP	Wkd	Wkr	EqP	Wkd	Wkr	EqP	
#5		x	x			x					x								
#8	x	x					x				x		x					x	
#13				x	x	x		x											x
#14	x	x																	

Hsh: household code; Wkd: increased proportion of weekend and holidays; Wrk: increased proportion of working days; EqP: approximately equivalent proportion between weekend and weekdays (eq. proportion is 71%:29%).

produce the loading identified in the analysis. It is clear that cross-referencing the load curves with the type of appliances used in the house and reported seasonal behavior would improve the interpretation of the load curves. It could help identify practices and "culprits" (i.e.: energy intensive appliances that mark load shape). An experiment is the consequence of an experimental design and of the questions planned in the beginning of the study, and the lack of resolution that would enable this cross reference analysis is a consequence of the initial design. Such analysis is an interesting gap to explore in future research.

6. Conclusions

Results show that it is possible to automatically (and under anonymity) extract and group persistent routine patterns in households. This information is useful as a way to discriminate consumption profiles in a population of a geographical area. Given the possibility that these patterns may change according to social-demographic characteristics such information is useful to help design better incentives (or other regulator mechanisms) for load shaping, and services that can be tailored to the characteristics of the populations.

The second main highlight of the research is the ability to study every individual family and cluster together groups of days that have similar baselines in common. Every day of the universe was identified by its loading characteristics, which are then interpreted once the days are clustered together into groups. The results are personalized baselines that have a certain probability to occur for any given day. Following the process described in the paper, for the majority of the households of the experiment, it is possible to identify the profiles that correspond to the following patterns:

1st baseline or the electricity consumed by vampire loads (stand-bys, off modes) and large appliances (refrigerators) without the presence of the occupants.
1st cold weekend days, which are profiles that occur in the winter over the weekend.
2nd cold working days. This baseline occurs mostly during the winter, and reflects the impact of electrical heating in the evenings.
2nd temperate days is a profile that can happen in any season and reflects the behavior of the occupants when there is no need for heating or air conditioning.

As put in evidence by the approach presented in this paper, pattern recognition analysis can be refined to characterize the response of households to specific conditions, given the intrinsic characteristics of the family, and it may be useful to help predict what families do next.

An alternative application of this approach is in the field of improving feedback provided to households about electricity consumption. In this case, comparisons of real time consumption could be made with the "best" consumption pattern of the recent–past for the same loading conditions. The development of this type of feedback mechanism could enable tailor-made suggestions that reflect the actual conditions of the household.

Acknowledgements

The authors acknowledge advice by Jorge Vasconcelos MIT Portugal Program and Inês Lima Azevedo, Carnegie Melon University. This research was supported by the Portuguese Science Foundation (SFRH/BD 40595/2007) and the German company YelloStrom, B.V.

References

[1] F. van Raaij, T. Verhallen, A behavioral model of residential energy use, Journal of Economic Psychology 3 (1983) 39–63.
[2] W. Heijs, Household energy consumption, in: W. Heijs (Ed.), Household Energy Consumption. Habitual Behavior and Technology, Springer, 2006.
[3] S. Darby, Smart metering: what potential for householder engagement? Building Research & Information 38 (October (5)) (2010) 442–457.
[4] F. Calabrese, J. Reades, C. Ratti, Eigenplaces: Segmenting Space through Digital Signatures, 2010.
[5] N. Eagle, A. (Sandy) Pentland, Reality mining: sensing complex social systems, Personal and Ubiquitous Computing 10 (November (4)) (2005) 255–268.
[6] O. Santin, Behavioral patterns and user profiles related to energy consumption for heating, Energy and Buildings 43 (2011) 2662–2672.
[7] Z. Yu, B. Fung, F. Haghighat, H. Yoshino, E. Morofsky, A systematic procedure to study the influence of occupant behavior on building energy consumption, Energy and Buildings 43 (2011) 1409–1417.
[8] I. Witten, E. Frank, Data Mining Practical Machine Learning Tools and Techniques, second ed., Morgan Kaufmann, San Francisco, 2005, p. 524.
[9] J. Maroco, Análise Estatística com o PASW Statistics. Pero Pinheiro: Report Number, 2010, p. 953.
[10] B. Moujalled, R. Cantin, G. Guarracino, Comparison of thermal comfort algorithms in naturally ventilated office buildings, Energy and Buildings 40 (12) (2008) 2215–2223.

Please cite this article in press as: J.M. Abreu, et al., Using pattern recognition to identify habitual behavior in residential electricity consumption, Energy Buildings (2012), doi:10.1016/j.enbuild.2012.02.044