Load profile disaggregation by Blind Source Separation: a Wavelets-assisted Independent Component Analysis Approach

Yongli Zhu
ECE Department, University of Tennessee, Knoxville
Knoxville, TN 37909, USA
zhuyongli@ieee.org

Songtao Lu
ECE Department, Iowa State University
Ames, IA 50011, USA
songtao@iastate.edu

***Abstract*— In this paper, a Blind-source separation method, i.e. Independent Component Analysis (ICA) is used for disaggregating the substation load profile into different patterns, i.e. residential and commercial groups. The smart meter data from a down town substation has been used. Principle Component Analysis (PCA) is applied for data reduction. Wavelet analysis is used to extract the trend signal from the original load profile as inputs for the ICA routine. Final results verify the effectiveness of this load profile disaggregation approach.**

***Index Terms*-- load profile; statistic learning; load modeling; smart meter; Blind-source separation**

I. INTRODUCTION

The load profile or load curve is widely used by power utilities to analyze consumers' usage pattern during a given period of time, which is useful in improving the load forecasting granularity by treating different user groups separately. They are valuable for both planning and operation of the power systems and for tariff design [1]. The identification and prediction of the load profile is not only a necessity for power generation and distribution companies, but also useful for small independent generation entity like owner of a microgrid in the future liberalized electricity market, because they can facilitate a more flexible demand response. In a word, the classified profiles will be helpful for [2-3]:

1) Planning: more refined load forecasting, i.e. forecasting residential, commercial and industrial loads separately.

2) Operation: more flexible energy dispatch strategy, e.g. arranging load shedding scheme according to the peak and valley time in the load profiles of different groups.

Previous related studies have mainly focused on two areas: 1) appliance monitoring and classification [4-8], which is another famous area, i.e. nonintrusive load monitoring (NILM); 2) daily load profile classification by large volume of historical data [9-15]. However, some shortcomings of those previous studies are:

1) Require *a priori* information: e.g. previous load profiles data for each group; the difficulty here is that in some situation for university researchers, the realistic load group data are not easy to obtain from utility side due to the confidentiality or commercial competition reasons; on the other hand, even utility itself might not have a great amount of these firsthand data because of the scarce deployment of smart meter or wave recorder in some less developed district due to cost consideration in upgrading legacy system.

2) Requires expensive hardware: switching event detectors at house level or advanced nonintrusive current sensors, e.g. magnetic or optoelectronics based.

3) Requires complex algorithm: finite state machine [4-6], fuzzy cluster [9], self organizing maps [10], spectral analysis [11], adaptive vector quantization [12], data mining [13], etc. In practice, those methods are not always computationally attractive especially when huge volume data flows in.

In terms of machine learning language, the previous study towards daily load profile classification problem when large amount of historical group profile data available can be classified as supervised learning type method ("supervised" can be understood as "having old data as reference and for training purpose"). In recent years, there also arise many unsupervised machine learning methods, among which some partially or totally depend on statistic technique and theory, i.e. so-called *statistic (machine) learning*. The Independent Component Analysis is a typical unsupervised statistic learning method, which was originally put forward as a *blind-source separation* method in radar signal processing application and is chosen as the foundation tool in this paper.

The succeeding sections are organized as follows: section II describes the some mathematical backgrounds for each of the three steps of the presented approach, i.e. PCA, Wavelets and ICA; section III shows the flowchart for the overall procedure; a case study and results analysis for a realistic down town substation have been carried out in section IV; Section V gives the conclusions and future study directions.

978-1-4799-6415-4/14/$31.00 ©2014 IEEE

II. METHODOLOGY

A. Data preprocessing and dimension reduction by PCA

For data preprocessing: the load profiles are examined in order to delete wrongly records. For PCA: the goal is to compute the most meaningful basis to re-express a noisy data set. It is desired that the new basis will filter outer the noise and reveal the hidden structure. In other words, PCA is to preserve the most significant basis vectors or more precise directions while discarding the redundant dimensions.

Suppose there are k signals $X = (X_1, X_2, \dots X_k), X_i \in R^n$. Then, the basic routine of PCA is as follows:

1) Find $Y = AX$, A is a transformation (projection indeed) matrix needs to be determined;

$$\begin{bmatrix} Y_1 \\ . \\ . \\ Y_k \end{bmatrix} = \begin{bmatrix} a_{11} & . & . & a_{1k} \\ . & & & . \\ . & & & . \\ a_{k1} & . & . & a_{kk} \end{bmatrix} \begin{bmatrix} X_1 \\ . \\ . \\ X_k \end{bmatrix} \tag{1}$$

2) The covariance matrix of Y is diagonal.

3) Select a subset of Y according to their covariance contributions (the first m largest ones, so reduced dimension from k to m).

B. Trend extraction by wavelets

Wavelet Transform provides an effective role for analysis of the component of data changing. The predominant advantage of wavelet analysis is the ability to perform local analysis which can reveal signal aspects that other analysis techniques miss, such as trends, through performing a multi-resolution analysis. The continuous wavelet transform is defined as:

$$x_{WT}(\tau, s) = \left(1 \big/ \sqrt{|s|}\right) \int_{-\infty}^{\infty} x(t) \psi^*((t-\tau)/s) dt \tag{2}$$

The transformed signal $x_{WT}(\tau, s)$ is a function of the translation parameter τ and the scale parameter s . The mother wavelet is denoted by ψ . A wavelet function has its own central frequency f_c at each scale, where the scale s is inversely proportional to that frequency. A large scale corresponds to a low frequency, giving global information of the signal. Small scales correspond to high frequencies, providing information in detail. Thus, wavelets have time-widths adapted to their frequencies. This is the main reason for the utilization of the wavelets in studying the changing of power load in trend by time-frequency signal analysis.

C. Load profile disaggregation by ICA

In recent research, ICA [16] is considered as a fairly effective statistical way due to its ability to separate a mixture of signal sources into different sources based on the signal component's high order statistical properties, as so called "blind-source separation". It has been successfully applied in signal processing area [17]. Essentially, it is an unsupervised statistic learning method. The main idea can be briefly expressed by the following mixed model:

$$x(t) = As(t) + n(t) \tag{3}$$

The statistical model in equation (3) is called ICA model, which describes how the observed data are mixed through the components $s(t)$. The m dimension column vector $x(t)$ is the observed data. A is a $m \times n$ mixing matrix; $n(t)$ denotes the additive noise vector. The matrix A is assumed to be unknown. All we observe is the random vector $x(t)$, and we must estimate both A and $s(t)$. Since A is unknown, so $s(t)$ seems to be unsolvable. Fortunately, there are many mathematical methods for calculating the coefficients of A by requiring the High-Order Statistics (HOS) information during the search for independent components. In this paper, the ICA is adopted for the disaggregation of load profile.

Besides, to check the rationality of the disaggregated signals in the sense of representing truly local residential and commercial profiles, we could refer to some typical customer profiles from utility reports or other literatures [9- 15].

III. FLOW CHART OF THE OVERALL PROCEDURE

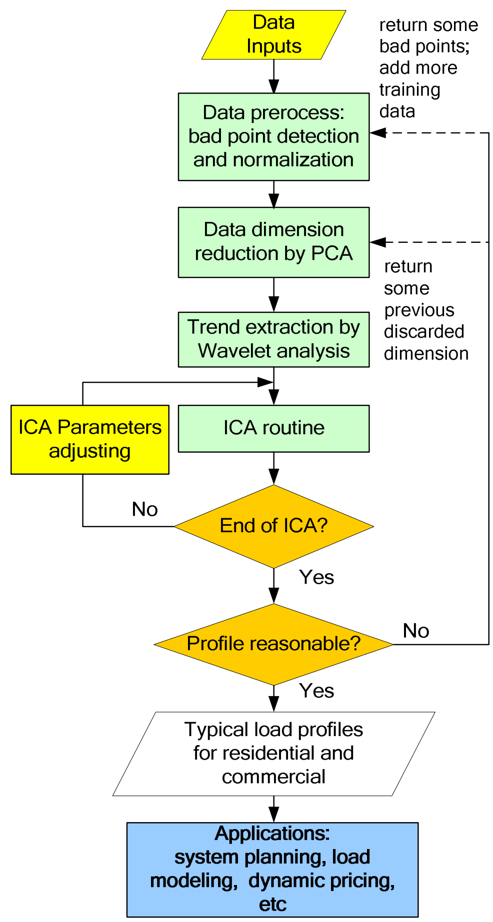


Figure 1. Flowchart of the overall procedure for load profile disaggregation

An overall flowchart of the presented approach is shown. Note that there are two auxiliary blocks, ICA parameter adjusting and profile rationality check. To some extents, the former depends on personal experience while the latter depends on the availability of reference sources like utility historical profiles database for different groups or existing results from other academic literatures (as used in this paper).

IV. Case Study

To demonstrate the application of this wavelet-assisted ICA approach for realistic data, the records made in a down town substation of local area have been used. The substation is located near the center of a medium sized town with a population of mostly students. Thus there are two main load groups, i.e. residential and commercial. In our case, the commercial section is in fact a mix of commercial and fewer industrial loads, thus it can be named as "quasi-commercial". In addition, all of the inhabitants are constant-community type, which means fairly low residential population mobility. That is to say, the substation raw data gives a representative load profile of the whole town. The data used is for the two summer months from May to July 2013. All measurements were taken at 15 minutes. There are totally 62 days in this period. In each daily profile, the time series ranges from 09:45am to 21:30pm. Note that the time unit in Fig. 2 is decimal day.

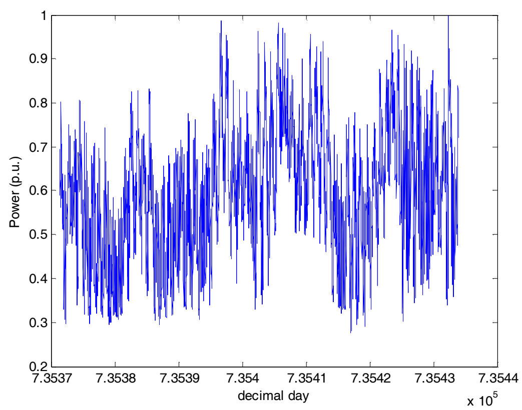


Figure 2. Original two months substation power profile

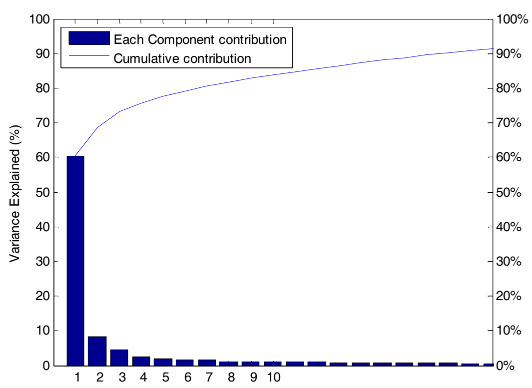


Figure 3. PCA result: percentage explained from each principle component

A. PCA results

The PCA result shows that the cumulative percentage contribution of the first 20 largest components exceeds 95% as shown in Fig. 3. And after the 61th component, each contributes zero. Compromise between dimension reduction and less loss of information is usually a challenge; while in our case, compared with the original size (96), 61 achieves both dimension reduction and minimum information loss. Thus in the following steps, these new data are utilized.

B. Wavelets results

Fig.4 shows the wavelet analysis result for a typical Saturday load signal by MATLAB wavelet tool box. The vertical axis means the signal magnitude (normalized between 0 to 1) and the horizontal axis shows the indices of the data samples (in this case it is 96 for one signal).It can be observed that the fourth layer (a4) signal can be used as the trend characteristics of the original signal. The layer number is chosen by trial and error as long as it provides enough information to reflect the accurate trend. Large layer number often means large computation cost, especially in utility application where rolling huge volume of load data on a daily basis is usually the case. In this case, the layer number is chosen as six. Other layers may also provide useful trend shape but with high frequency noise or components which are not desired in the next step as inputs to the ICA routine.

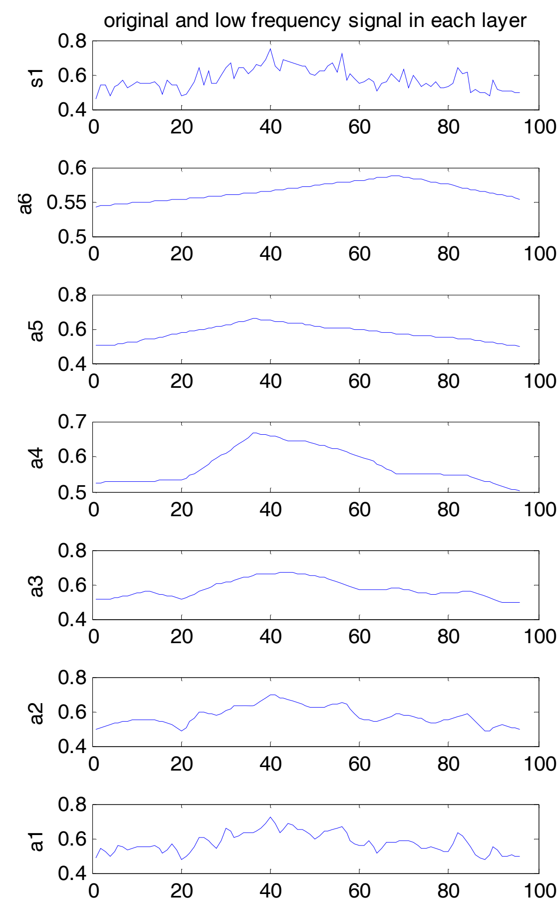
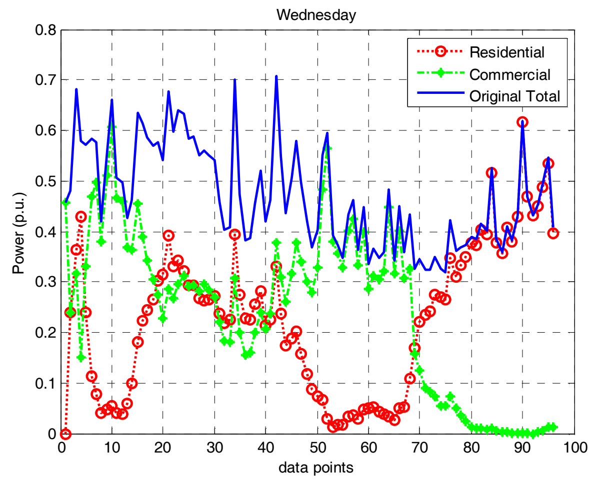


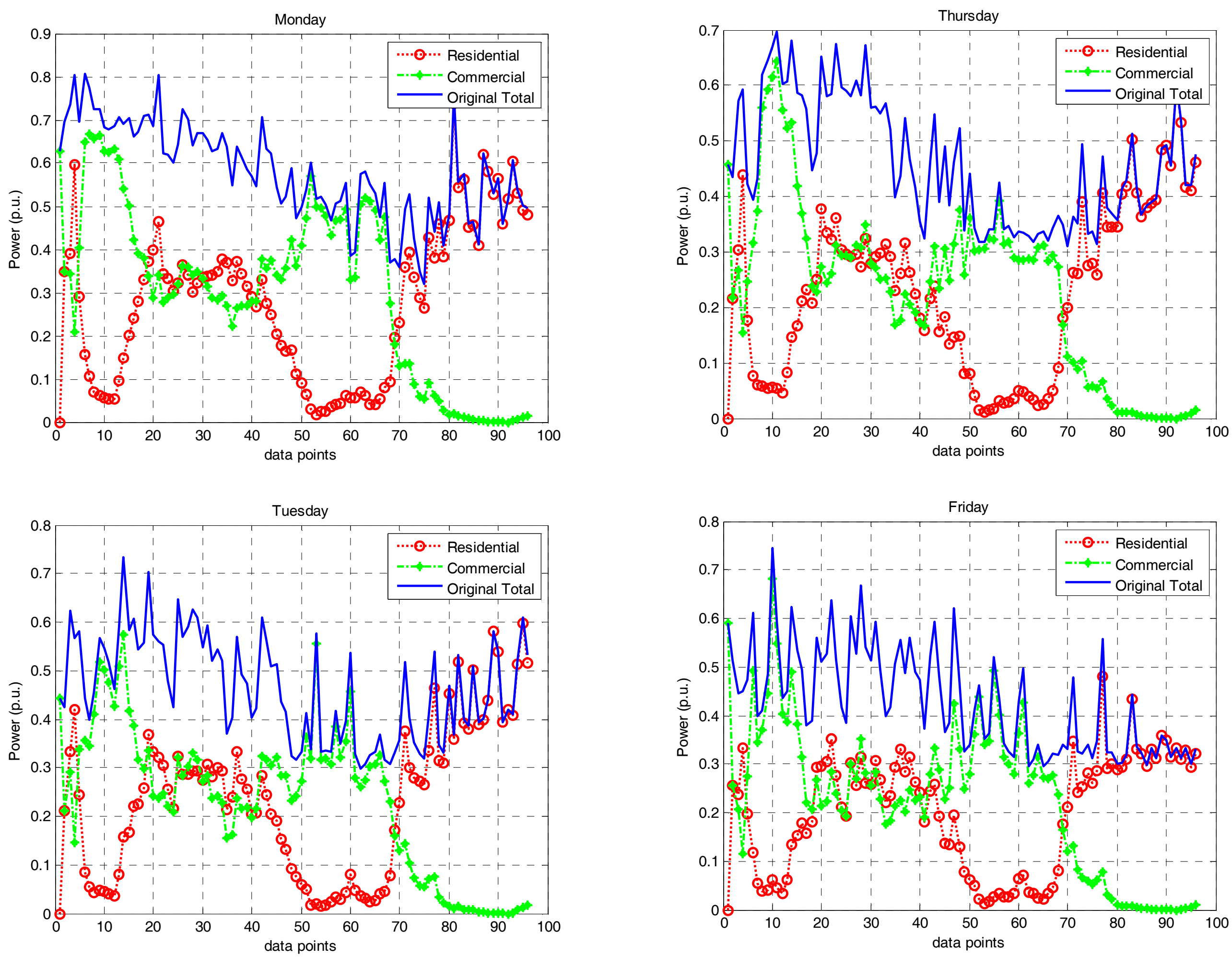
Figure 4. Six layers wavelet decomposition: low frequency component (s1: original signal; a1 to a6: low frequency component)

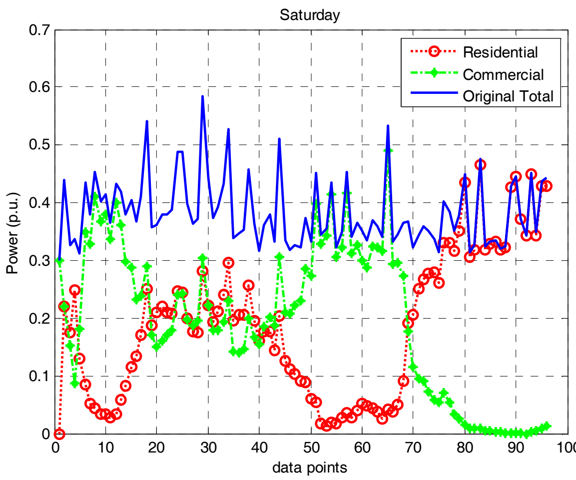
In other words, the different groups are distinct to each other by their corresponding low-frequency traits, e.g. for commercial and residential patterns, they are different over the whole 24hr; but in some instant, they can be similar.

C. ICA results

By the FastICA algorithm, the ICA disaggregation results are shown in Fig. 5. Here, the first week load profiles are selected to display with 96 points in one day (horizontal-axis). The plots indicate that the residential group has one peak period around 12pm with another one around 7:30pm to 2:30am; while commercial group has three peak periods: a relatively flat period ranges from 09:30am to 11:30am, then 12pm to 10pm and finally 5am to 8am. Note that the two patterns are nearly complementary; this can be explained by the phase separation nature of ICA and the night operation of some commercial loads like pubs and office equipments







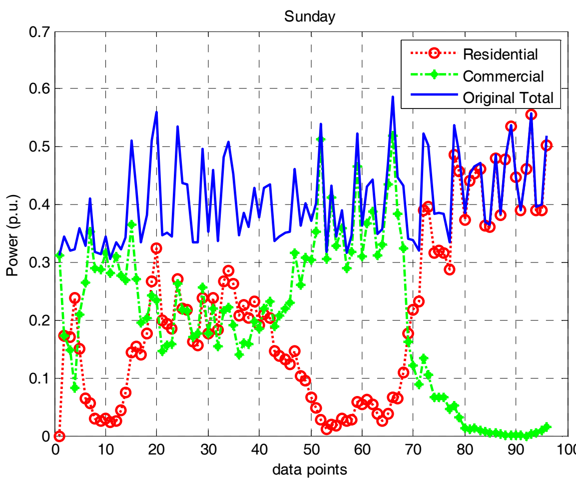


Figure 5. Load profile disaggregation by ICA: from Monday to Sunday in the first week (Y-axis: Power p.u.; X-axis: 96 data points in one day)

V. CONCLUSION

This paper investigates a methodology of disaggregating the total load profile from substation. The blind-source separation method ICA is applied based on unsupervised statistic learning idea. Then Wavelet analysis and PCA is used for dimension reduction such that the principle properties of power load changing can be characterized within a relatively small subspace of the original data.

Finally the representative daily load profiles are approximately obtained. This information is helpful for power utilities in load forecasting, load management, and tariff design. By the disaggregation results, the system operator can adaptively make the power dispatching decision by a more updated peak-load time curves for different load groups.

Another advantage of the method show that it only need the load profile at substation level rather than collecting recorder data at each different group sites, so the communication load and computational burden can be reduced significantly.

For future study, some directions can be:

1) Refined short term load forecasting by different groups utilizing the disaggregation (classification) results from ICA;

2) Considering other significant factors like temperature, season, day type (weekend and weekdays) to achieve better disaggregation effect.

3) Modify the ICA algorithm to make it more specialized for load profile decomposition problem.

REFERENCES

[1] G. Chicco, R. Napoli, P. Postolache, M. Scutariu and C. Toader. Customer characterization options for improving the tariff order. *IEEE Trans. Power Syst*., vol. 18, no. 1, pp. 381 – 387, February 2003.

[2] J. Jardini, C. Tahan, M.R. Gouvea, et al. Daily load profiles for Residential, Commercial and Industrial Low Voltage Consumers. *IEEE Trans. Power Delivery*, vol. 15, no. 1, pp. 375 – 380, January 2000.

[3] D. Gerbec, S. Gasperic, F. Gubina. Determination and allocation of typical load profiles to the eligible consumers. IEEE Power Tech Conference, Boloogna, Italy, June 2003.

[4] Cox, R.; Leeb, S.B.; Shaw, S.R.; Norford, L.K. Transient event detection for nonintrusive load monitoring and demand side management using voltage distortion. Applied Power Electronics Conference and Exposition, 2006. APEC '06.

[5] Ruzzelli, A.G.; Nicolas, C.; Schoofs, A.; O'Hare, G. M P. Real-Time Recognition and Profiling of Appliances through a Single Electricity Sensor. Sensor Mesh and Ad Hoc Communications and Networks (SECON), 2010 7th Annual IEEE Communications Society Conference on , vol., no., pp.1,9, 21-25 June 2010.

[6] Froehlich, J.; Larson, E.; Gupta, S.; Cohn, G.; Reynolds, M.S.; Patel, S.N. Disaggregated End-Use Energy Sensing for the Smart Grid. *IEEE Pervasive Computing*, vol.10, no.1, pp.28, 39, Jan.-March 2011.

[7] C. Laughman, K. Lee, R. Cox, et al. Power signature analysis. *IEEE Power and Energy Magazine*, vol.1, no.2, pp.56,63, Mar-Apr 2003

[8] Weiss, M.; Helfenstein, A.; Mattern, F.; Staake, T. Leveraging smart meter data to recognize home appliances. 2012 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp.190, 197, 19-23 March 2012

[9] G. Chicco, R. Napoli and F. Piglione. Comparison Among Clustering Techniques for Electricity Customer Classification. *IEEE Trans. Power Syst*., vol. 21, no. 2, pp. 933 – 940, May 2006.

[10] G. Chicco, R. Napoli, F. Piglione. Application of clustering algorithm and self organizing maps to classify electricity customers, Power Tech Conference Proceedings, 2003 IEEE Bologna, vol. 1, pp. 7, June 2003.

[11] E. Carpaneto, G. Chicco, R. Napoli, M. Scutariou. Electricity customer classification using frequency-domain load pattern data. *Electrical Power and Energy Syst*. vol. 28, no. 1, pp. 13-20, 2006.

[12] G. Tsekouras, N. Hatziargyriou and E Dialynas. Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers. . *IEEE Trans. Power Syst*., vol. 22, no. 3, pp. 1120-1128, August 2007.

[13] V. Figueiredo, F. Rodrigues, Z. Vale, J. B. Gouveia. An electric energy consumer characterization framework based on data mining techniques. *IEEE Trans. Power Syst*., vol. 20, no. 2, pp. 596-602, May 2005.

[14] A. Mutanen, M. Ruska, S. Repo. And P. Järventausta. Customer Classification and Load Profiling method for Distribution Systems. *IEEE Trans. Power Delivery*, vol. 26, no. 3, pp. 1755 – 1763, July 2011.

[15] H. Liao, R. Napoli and D. Niebur. Load profile Estimation in Electric Transmission Network Using Independent Component Analysis. *IEEE Trans. Power Syst*., vol. 18, no. 2, pp. 707 – 715, May 2003.

[16] A. HyvAarien. A family of fixed-point algorithms for independent component analysis. Proceedings of IEEE International conference on Acoustics, Speech and Signal Processing (ICASSP'97), Munich, Germany, 1997, pp.3917-3920.

[17] S. Lu, Y. Zhao. A multipath mitigation algorithm in Global Navigation Satellite Systems arrays using Independent Component Analysis. 2010 IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS), Beijing, pp.28-32, 25-27 June 2010.