Disaggregating and Forecasting the Total Energy Consumption of a Building: A Case Study of a High Cooling Demand Facility

Juliana Barcelos Cordeiro, Khashayar Mahani, Farbod Farzan, Mohsen A. Jafari

Abstract**—**Energy disaggregation has been focused by many energy companies since energy efficiency can be achieved when the breakdown of energy consumption is known. Companies have been investing in technologies to come up with software and/or hardware solutions that can provide this type of information to the consumer. On the other hand, not all people can afford to have these technologies. Therefore, in this paper, we present a methodology for breaking down the aggregate consumption and identifying the high-demanding end-uses profiles. These energy profiles will be used to build the forecast model for optimal control purpose. A facility with high cooling load is used as an illustrative case study to demonstrate the results of proposed methodology. We apply a high level energy disaggregation through a pattern recognition approach in order to extract the consumption profile of its rooftop packaged units (RTUs) and present a forecast model for the energy consumption.

Keywords**—**Energy consumption forecasting, energy efficiency, load disaggregation, pattern recognition approach.

I. Introduction

IN 2014, residential and commercial buildings consumed 41% of energy resources in the US [1]. Studies have shown that energy efficiency can be improved by 5 to 15% when it is known which devices are consuming the energy [2], and this information can be achieved by breaking down the energy load, also known as energy disaggregation. Thus, due to the importance of load disaggregation, many companies such as Bidgely, Enetics and Intel, have been developing non-intrusive load monitoring (NILM) technologies to break out energy loads [3].

Appliance Load Monitoring (ALM) methods can be found in the literature as methods that perform energy sensing and provide the breakdown of the energy data. They have NILM and ILM as their major approaches. The NILM approach only requires a single meter per building or residence and it was developed as a cheaper alternative to the ILM since this one requires at least one sensor per appliance [4].

According to Hart [5], "A nonintrusive appliance load monitor determines the energy consumption of individual appliances turning on and off in an electric load, based on detailed analysis of the current and voltage of the total load, as measured at the interface to the power source". Since not all appliances consume constant power, Hart [5] proposed following operational state groups:

1. On/Off appliances: this group comprehends most of the household appliances that at any given time may be either on or off, but with only a single type of ON state, for example light bulb and water pump;
2. Finite State Machine (FSM): multi-state appliances with finite operation states belong to this group, such as washing machine and dishwashers;
3. Continuously Variable: this group comprehends the appliances with infinite operation states, such as light dimmers and sewing machines.

A fourth operational state group is defined in [6] as "Permanent consumer devices" which includes appliances that are on ON mode 24 hours per day and 7 days in the week and have their active and reactive power consumption approximately constant, such as aquarium pump.

The appliance features used to disaggregate the energy consumption through NILM approaches can be categorized as steady state (power change features, V-I trajectory, voltage noise, etc.), transient state (transient power, start-up current waveforms, voltage noise) and non-traditional [4].

Even though NILM techniques are powerful, they are impracticable when dealing with low-frequency data (1h – 15min), wherefore some authors use pattern recognition approaches to estimate high level properties from the energy consumption [7]-[9].

In this work, we applied the pattern recognition methodology. Resolution of utilized time-series data is 15 minutes and no information about appliances features is recorded.

One may not only want to know which appliances are impacting the total energy consumption but also predict how much the total consumption is going to be due to changes in certain controllable input, for instance asset operation schedule. Combining energy disaggregation method with prediction of energy consumption in end uses will create the beneficial approach to control the building energy consumption and improve the energy efficiency. Once the user can forecast the energy consumption, the mathematical programming could be formulated to minimize the energy consumption, which is an advantage for both consumer and the environment. However, the energy system of building is

J. B. Cordeiro is with the Industrial and Systems Engineering Department, Rutgers University, Piscataway, NJ 08904 USA (phone: 848-203-8120; e-mail: juliana.cordeiro@rutgers.edu).

K. Mahani and M. A. Jafari are with the Industrial and Systems Engineering Department, Rutgers University, Piscataway, NJ 08904 USA (e-mail: km723@scarletmail.rutgers.edu, jafari@rci.rutgers.edu).

F. Farzan is with the Civil and Environmental Engineering Department, Rutgers University, Piscataway, NJ 08904 USA (e-mail: farbod@scarletmail.rutgers.edu).

quite complex since different energy types can compose the system, such as heating/cooling load, electricity and hot water. The building energy behavior is dependent on factors like occupancy, weather conditions, HVAC (Heating, Ventilating and Air-Conditioning) systems and their schedules, etc. Therefore, due to the complexity of the problem, a precise forecasting is difficult.

The energy consumption prediction methods can be divided as engineering method, statistical methods, neural networks, support vector machines and grey models [10], [11]. The one chose to perform this study was the statistical method, which correlates the energy consumption to the variables of influence.

In conclusion, in this paper, we perform a methodology for breaking down the aggregate power consumption into the main end-uses profiles and creating the forecast model for these energy consumers. The load-forecasting model could be used for optimal operational control of any kind of building. We demonstrate a performance of such approach for an illustrative case study, an establishment that requires high cooling. This establishment will be denominated here as "Facility X".

The paper is structured as follows. In Section II, a general problem definition is presented while in Section III, the case study problem is described. Section IV describes the methodology utilized to extract RTUs profile from aggregate metered data. In Section V, preliminary data analysis is demonstrated to show the correlation between consumption data and some important measurable features such as ambient temperature and internal temperature. In Section VI, the regression model, used to forecast the consumption profile of RTUs will be explained and finally the conclusion is presented in Section VII.

II. Problem Definition

Even though techniques such as NILM have been focused by many authors in order to perform energy disaggregation, they turn to be more difficult to perform when the data set is not complete. For example, one may pursue data to analyze in which the time horizon does not fit on the NILM techniques or even do not have enough information about the existents appliances on the building in study, which is a challenge when analyzing the data.

In this work, we use pattern recognition approach to break down the aggregate consumption in its major end-uses profiles. This is done by observing if the total power consumption varies when the operational state (on/off) of a particular appliance is changed.

After knowing the major energy types that compose the energy system of a building, it is possible to estimate the energy consumption through a forecasting model.

A. Forecasting Formulation

Since building energy behavior depends on many factors, in this paper we considered features such as schedule of the RTUs, external and internal temperature, and the time of the day as independent variables of our linear regression model which has the sub-hourly energy consumption (kWh) as the dependent response variable. The energy consumption model proposed in this work is adapted from [12] and, given n sub hourly intervals, it can be written as:

$$\widehat{Y}_i = \beta_0 + \beta_1 x_{1i} + \sum_{k=1}^{5} \beta_{2k} x_{2ik} + \sum_{k=1}^{5} \beta_{3k} x_{3ik} + \sum_{k=1}^{5} \beta_{4k} x_{4ik} + \sum_{m=1}^{95} \beta_{5m} x_m \quad \forall \quad i = 1, 2, \dots n \tag{1}$$

where $\widehat{Y}_i$ is the estimated energy consumption; all β are the parameter estimators; x_{1i} is the total number of RTUs operating; x_{2ik} is the difference between the internal and external temperature for each zone k; x_{3ik} is the difference between the zone temperature at time t and at time $t-1$ in zone k; x_{4ik} is the cooling set-point temperature of RTU in zone k; and x_m is a dummy variable to indicate time interval of the day.

III. Case Study Description

The Facility X has 6 RTUs and two energy meters that collect the total energy consumption of the building.

The available range data to perform this study is 8/19/2009 to 11/08/2012 and has two time scales:

- 15 minutes resolution: Energy consumption data for both meters, zone temperature, set-point temperature and state of operation for each RTU;
- Daily resolution: Energy consumption data and average outside temperature.

The energy consumption data for each RTU is not available and it is not known each meter is connected to which assets. Thus, in order to extract the consumption profile of the RTUs it is necessary to analyze the consumption data and schedule of the Facility X to disaggregate the metered consumption data. And, as second step, in order to forecast the energy consumption, it is necessary to analyze the temperature data, RTUs schedule and energy consumption.

IV. Disaggregation Methods and Results

Since the RTUs are the only appliance data available for Facility X, in our study we applied a pattern recognition approach in order to identify which meter collect data from those assets and if the RTUs are cooling or heating systems.

As mentioned earlier, energy consumption data for both meters with daily and 15 minutes time resolution, and the states of operation (occupied/setback) of each RTU with 15 minutes resolution are provided.

The method used in this study to disaggregate the meters data was by recognizing the impact caused on the pattern of the power consumption caused by the RTUs' schedule. In other words, if the power consumption varies when the RTU state of operation changes it means that the meter is responsible for collecting the respective RTU energy consumption.

In Figs. 1 (a) and (b), power consumption and schedules for 08/19/2009 are demonstrated, respectively. It can be seen that the power consumption of Meter 1 starts by the time the RTUs

1, 3 and 5 have their operation state changed from setback to occupied, and a slightly increase in the power consumption occurs when RTUs 2 and 4 have their operation state changed as well. The same pattern can be observed in the end of the day: when RTUs 2 and 4 turn to the setback point, the power consumption has a visible drop and then goes back to zero, which can be explained by the change in the operation state of the RTUs 1, 3 and 5. In respect to RTU 6, operation state changes results in Meter 2 profile changes and doesn't have impact on Meter 1 consumption profile.

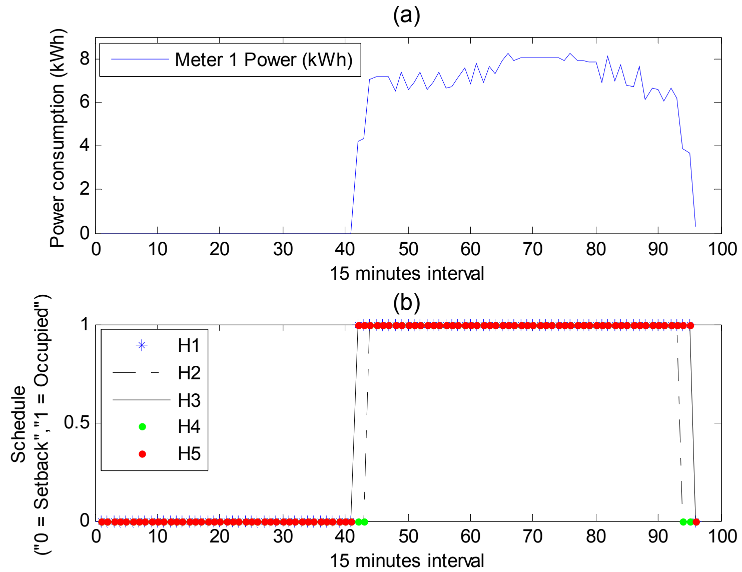

Fig. 1 (a) Meter 1 Electricity Consumption and (b) RTUs' schedule strategy in 08/19/2009

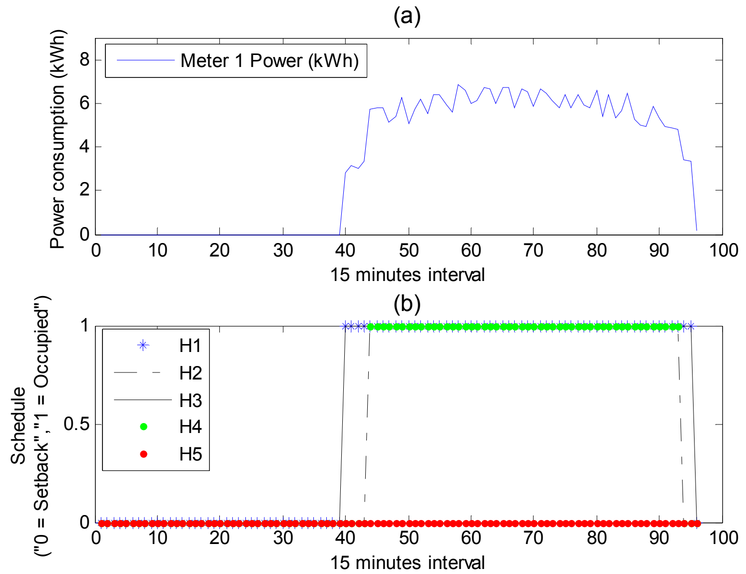

Fig. 2 (a) Meter 1 Electricity Consumption and (b) RTUs' schedule strategy in 08/19/2010

Figs. 2 (a) and (b) show the power consumption and the RTUs' schedule, respectively, of the following year (08/19/2010), where it can be seen that a drop in the maximum power consumption of Meter 1 occurs (approximately from 8kWh to 7kWh) as the same time RTU 2 has its operation state on setback point during the entire day.

The following step is to analyze if the 5 RTUs load designated to Meter 1 are cooling and/or heating. Therefore, in order to analyze it, a comparison between Meter 1 power consumption and average zone temperature of the 5 RTUs was performed and a sample of the summer (August) can be seen in Fig. 3. It can be seen that the average internal temperature is high during the time that the power consumption is zero, and it starts to decrease as power consumption increases and vice-versa. For winter samples, a heating pattern was observed since the average zone temperature is low when the energy consumption is 0 and while the energy consumption increases the temperature gets higher values.

In conclusion, RTUs 1, 2, 3, 4 and 5 are responsible for the store cooling and heating and have Meter 1 collecting their energy consumption, while Meter 2 collects the energy consumption of RTU 6, general lighting and equipment.

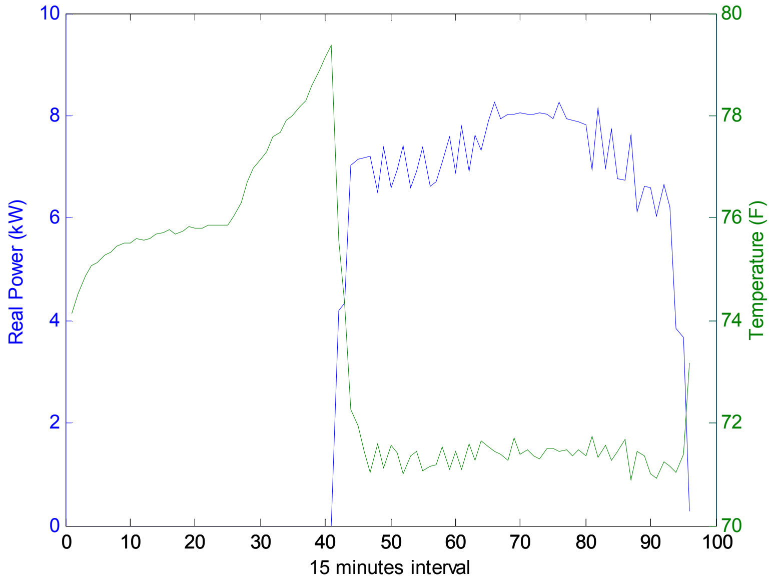

Fig. 3 Meter 1 Electricity Power x Internal Temperature

V. Preliminary Data Analysis

In this section, we are going to perform feature selection analysis in order to find the variables that are correlated with the energy consumption, such as temperature data and RTUs schedule, for a daily level.

The features' combinations are:

- A–Average outside temperature;
- B–RTUs' hours of operation and average outside temperature;
- C–Average zone temperature, RTUs' hours of operation and average outside temperature.

The Regression statistics results, for the three combinations groups, can be seen on Table I. The *R-Square* values show that for every feature added, the dependence of Meter 1 is increased.

TABLE I
Regression Statistics for A, B, and C combinations

Statistics	A	B	C
R-square	0.846	0.876	0.902
Adjusted R-square	0.846	0.875	0.900
Standard Error	54.00	48.46	43.22
Observations	994	994	994

Finally, we conclude that Outside Temperature, RTU's schedule strategy and Zone temperature are important features that must be included in the multiple linear regression model, which is develop the following section.

VI. Forecast Model

The energy consumption prediction method applied in this paper is the multiple linear regression in which the independent variables considered are: Total number of RTUs operating, difference between RTU's respective zone temperature and external temperature, difference between zone temperature at time t and time $t-1$ in each control zone, RTU's cooling set-point temperature, and the dummy variable indicating the time interval of the day, and the dependent variable is the sub-hourly electricity consumption of Meter 1 (kWh).

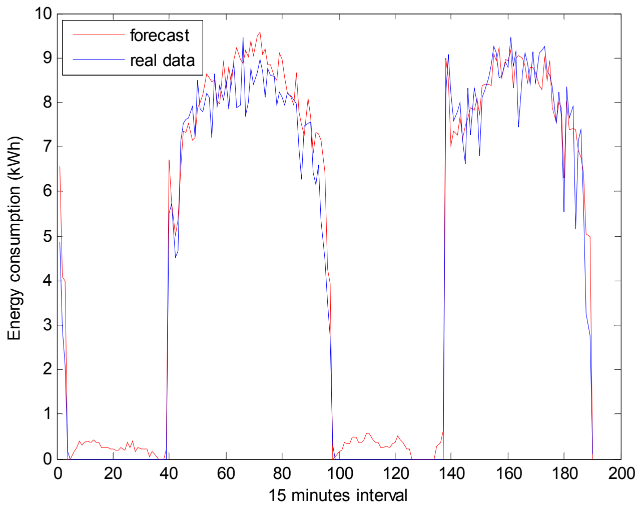

Fig. 4 Real Energy Consumption vs. Forecasting

The forecast model developed in this paper will focus on a cluster composed by 5 months with highest consumption level namely; May, June, July, August and September.

We used hold out methodology to measure the error of the forecast model. Meaning, we split the data into two groups: the training and the testing sets. As mentioned in section 3, our dataset is from 8/19/2009 to 11/08/2012. We considered 2 summers as training set and the other summer as the testing set.

In order to develop the linear regression and obtain an equation such as in (1), we considered 2 sets of data as training data and then, to validate the model, we tested it on a 3rd set of data.

The multiple linear regression was performed in MATLAB and, as it can be seen on Table II, the forecasting model developed has a root mean square error (RMSE) of 1.07 and an R-square value of 0.865, which shows that the independent variables considered in proposed model have great impact on the energy consumption (dependent variable). Fig. 4 illustrates the comparison between the real energy consumption and the forecasted values for two sample days (July 10th and July 11th).

Upon completion of the training step and obtaining the multiple linear regression model, we evaluated the performance by analyzing the real prediction error of the model through its application to the testing set. As result, the RSME for testing set is 1.2345 kWh which is acceptable since the peak of consumption is around 9.5 kWh.

TABLE II
REGRESSION STATISTICS FOR FORECASTING MODEL

Statistics	Value
R-square	0.865
Adjusted R-square	0.865
Root Mean Square Error	1.07
Observations	29193

VII. CONCLUSION

Building energy efficiency has been on focus since it represents a big percentage of the total energy consumption of a region. Energy efficiency brings advantages for both the environment and the consumer, wherefore companies have been investing on technologies that help the user to save energy and, as result, money. On the other hand, not all consumers can afford such technologies, even though companies promise they will pay back over time.

Studies have shown that when the consumer know which equipment/assets impact in their building's energy consumption, energy efficiency can be achieved by up to 15%. Companies have been investing on NILM technologies to perform energy disaggregation, but since these techniques are impracticable when dealing with low-frequency data (1h – 15min) an alternative is to use pattern recognition approach, which estimates high level properties from the energy consumption.

Another way of saving energy is by predicting the energy consumption due to certain features such as schedules and changing on the weather conditions, for example. Multiple linear regression is the statistical method used in this work, which is among many methods to forecast energy consumption.

In conclusion, in this paper we combine the two major energy efficiency techniques and show our methodology through a case study. The building in study is a generic facility categorized by high cooling demand in which the data given was limited to aggregate meter energy data, rooftop package units (RTUs) schedules and features such as cooling and heating set-points. Besides not having more information about the building and its assets, the facility has two meters and no information about their energy collection division was known.

Finally, considering that the facility fits in the group of users that cannot afford energy saving technologies, we performed a pattern recognition approach to identify the major end-uses profiles of its energy consumption so that an energy forecasting model could be developed. The multiple linear regression performed for the training set has an R-square value of 0.865 and the root square mean error (RSME) for the testing set is acceptable for a value of 1.2345 kWh since the peak of consumption is around 9.5 kWh.

REFERENCES

[1] U.S. Energy Information Administration. "How much energy is consumed in residential and commercial buildings in the United States?" Available at: http://www.eia.gov/tools/faqs/faq.cfm?id=86&t=1

[2] S. Darby, "The effectiveness of feedback on energy consumption." Environmental Change Institute, University of Oxford, 2006. Available at: http://www.globalwarmingisreal.com/energyconsump-feedback.pdf. Visited: September 2015

[3] J. S. John, "Putting energy disaggregation tech to the test," November, 2013. Greentech Media. Available at: http://www.greentechmedia.com/articles/read/putting-energy-disaggregation-tech-to-the-test. Visited: September 2015

[4] A. Zoha, A. Gluhak, M. A. Imran, S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey," *Sensors*, vol. 12, no. 12, pp. 16838-16866, December 2012.

[5] G. W. Hart, "Nonintrusive appliance load monitoring," in *Proc. of the IEEE*, vol. 80, pp. 1870-1891, December 1992.

[6] M. Baranski, J. Voss, "Non-intrusive appliance load monitoring based on Optical Sensor," *IEEE Bologna PowerTech Conference*, Bologna, Italy, June 2003. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1304732

[7] L. Farinaccio, R. Zmeureanu, "Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major en-uses," *Elsevier*, Energy and Buildings, vol. 30, no. 3, pp. 245-259, August 1999.

[8] J. M. Abreu, F. C. Pereira, P. Ferrão, "Using pattern recognition to identify habitual behavior in residential electricity consumption," *Elsevier*, Energy and Buildings, vol. 49, pp. 479-487, June 2012.

[9] C. Beckel, L. Sadamori, S. Santini, "Automatic socio-economic classification of households using electricity consumption data," in *Proc. of the 4th international conference on future energy systems*, New York, 2013, pp. 75-86.

[10] H. Zhao, F. Magoulès, "A review on the prediction of building energy consumption," *Elsevier*, Renewable and Sustainable Energy Reviews, vol. 16, no. 6, pp. 3586-3592, August 2012.

[11] G. K. F. Tso, K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Elsevier*, Energy, vol. 32, no. 9, pp. 1761-1768, September 2007.

[12] F. Farzan, S. A. Vaghefi, K. Mahani, M. A. Jafari, J. Gong, "Operational planning for multi-building portfolio in an uncertain energy market," *Elsevier*, Energy and Buildings, vol. 103, pp. 271-283, September 2015.