ELSEVIER

Energy and Buildings 30 (1999) 245–259

ENERGY AND BUILDINGS

www.elsevier.com/locate/enbuild

Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses

Linda Farinaccio, Radu Zmeureanu *

Department of Building, Civil and Environmental Engineering, Centre for Building Studies, Concordia University, Montreal, Quebec, Canada H3G 1M8

Received 30 October 1998; accepted 2 February 1999

Abstract

The method presented in this paper shows a promising potential for application in residential buildings. The results prove that the whole-house electricity consumption can be disaggregated into its major end-uses, using a pattern recognition approach and only one sensor installed on the main electric entrance of the house. It also required a one-time submetering of the target appliances during the training period, of about a week, to find the electric characteristics of appliances. The results are provided in terms of daily load profiles, energy consumption and energy contribution of selected appliances. The proposed method was tested with monitored data from 3 weeks: (i) the training period of 1 week in October, (ii) the near-to-date testing period of 1 week in November and (iii) the far-to-date testing period of 1 week in January. For instance, the difference between monitored and estimated contribution is, for the month of October 1996, as follows: (i) 13 kW h or \$0.85 for the DHW heater and (ii) 6 kW h or \$0.36 for the refrigerator. The overall difference for both appliances does not exceed \$1.25 for the month of October, for a total electricity bill of 912 kW h and \$60.60, which appears to be acceptable for every homeowner. The errors in evaluating the daily energy consumption is between −10.5% and 15.9% for both the DWH heater and the refrigerator. © 1999 Elsevier Science S.A. All rights reserved.

Keywords: Residential buildings; Energy; Equipment; Monitoring; Field; Pattern recognition

1. Introduction

The breakdown of whole-house electricity consumption among the major end-uses is beneficial to increase the homeowners' awareness about the actual energy performance of houses. Traditional load monitoring techniques can be described as intrusive techniques due to the physical placement of sensors on individual appliances to gather end-use load data. Thus, this poses as a long-term intrusion onto the private life and property. More recently, researchers have developed non-intrusive techniques of load monitoring as an alternative to long-term intrusive metering. Non-intrusive techniques of load monitoring are based on the analysis of appliance energy signatures. An appliance signature gives information about the operating state of an individual appliance, using the monitored whole-house electric demand. The main advantage of defining appliance signatures in terms of the whole-house load is that, afterwards, only a single monitoring point in the house is required to gather end-use load data. The appliance signature, like the building signature, is assumed to remain constant for the life of the appliance given that no modifications are made or malfunctions occur.

The simplest appliance signature is defined as a two-dimensional signature vector representing the step changes in the measured power of an appliance. Power is just one type of signature, but in fact, appliance energy signatures can be defined in several other ways [1]. Most common residential appliances can be modeled using steady-state signatures, which have the following main advantages: (i) they provide a continuous indication of the appliance's operating state, making it much easier to detect the change; (ii) the sum of the power changes in any cycle of state transitions is zero; and (iii) they are additive when two appliances are activated.

Two existing systems for non-intrusive load monitoring are shortly presented below: (i) the Non-Intrusive Appliance Load Monitoring (NALM), developed at the Massachusetts Institute of Technology and (ii) the Heuristic End-Use Load Profiler, developed by Quantum Consulting, in California.

* Corresponding author. Tel.: +1-514-848-3200; fax: +1-514-848-7965; e-mail: zmeur@cbs-engr.concordia.ca

0378-7788/99/\$ - see front matter
PII: S0378-7788(99)00007-9

1.1. Non-intrusive appliance load monitor

The non-intrusive appliance load monitor is a device used to sample the total electrical demand at a high speed and to analyze the collected information in order to determine when appliances switch ON and OFF [1–3]. The prototypical NALM for residential buildings has five steps of operation: (1) measurement of current and voltage using a 1-s sampling interval, from which real and reactive power are calculated, (2) detection of ON and OFF events, (3) clustering of similar events, (4) matching of ON and OFF events over time and (5) equipment identification. When an appliance switches ON or OFF, the power level changes and a new steady power level is established. The difference between the two steady power levels defines an event. Appliance identification is then made by comparing the event power with known characteristics of typical appliances. Results indicate that energy consumption estimated by NALM is usually within 10% of sub-metered values. The results for stoves typically yield higher errors due to many simultaneous events that can go undetected. Another possible source of error is due to incorrect installation of the device.

Other researchers from Japan and the United States have applied similar techniques to the residential gas appliances [4]. The equipment used for this purpose consists of a Smart Gas Meter equipped with a pulse meter, a data logger and a software. The gas meter translates the movement of the diaphragm into a cyclic crank rotation and then transmits it to a digital indicator. One crank rotation is set equal to the minimum value of mechanically distinguishable unit of flow. A magnet attached to the gas meter then emits an electronic pulse as the crank rotates. Another device detects the pulse and is connected to a microcomputer in the Smart Gas Meter. Given the volume of gas that flows between two pulses through a standard size meter, and a set of heuristic rules about the flow rate and duration of various gas appliances (based on house-specific audit information), the use of individual appliances can be detected. The results indicate the method is about 95% accurate.

1.2. Heuristic end-use load profiler

The Heuristic End-Use Load Profiler (HELP) is a PC-based proprietary software package consisting of a rule-based disaggregation algorithm [5–7]. The software, used by approximately 30 to 40 users, is mainly oriented towards electric utilities [8]. The algorithm uses measurements of premise-level data, appliance connected loads and customer appliance ownership data, along with assumptions about the customer behavior, in order to construct the estimated end-use load profiles. In cases where two or more major end-uses have similar demand levels, the algorithm uses decision analysis techniques to distinguish between them, based on the assumptions regarding the usage of these appliances, such as the time of day or the length of usage. HELP has been used to produce load profiles for residential air conditioning units, HVAC equipment and water heating tanks. For instance, in a study of 40 houses over 4 months, the average air conditioner energy consumption was estimated within 10% of the actual energy consumption.

This paper presents the development of a new rule-based pattern recognition approach, used to disaggregate the total electricity consumption of a house into the major end-uses. The objective of the work presented herein is to demonstrate that end-use load data can be obtained by using only the whole-house load data, and applying a pattern recognition approach, to detect individual appliance loads from rapid sampling of electric current at the main entry point into the house. The results are provided in terms of daily load profiles, energy consumption and energy share of some selected appliances.

The recognition of the various energy demand patterns in an existing house consists of two phases: (i) a one-time training process required to tune the generic algorithms, previously developed, to the characteristics of electric demand of each particular house under investigation and (ii) the application phase. During the training phase, the electric current is monitored at the main electric entrance of the house as well as at several appliances of interest such as the domestic hot water tank or the refrigerator. In order for the approach to be cost effective, the training phase should have a minimum duration, for instance 1 week. In the application phase, only the electric current at the main electric entrance is monitored.

This paper presents (i) the development of the generic algorithms, (ii) the validation of algorithms, by comparing their results with the measurements performed at the selected appliances and (iii) a sensitivity analysis to evaluate the usefulness of all proposed rules.

2. Pattern recognition approach

The new approach is based on the top-down mental process of recognizing an object [9]. In this paper an object is the shape of variation of electric current or demand, monitored at the main electric entrance of the house, during the operation of a particular appliance. In the first step, the pre-existing knowledge, collected during the training phase, is used to generate a hypothesis about the object (e.g., the increase of electric current or demand is due to the activation of a refrigerator). The step-increase or decrease of the electric current or demand (called herein ‘initial signal’), due to the activation or deactivation of an appliance, is a distinct feature which can be used for the preliminary recognition of objects. However, the hypothesis could be false for at least two reasons: (i) the signal is generated by the simultaneous operation of other appliances and not by a single appliance and (ii) there are two or more appliances which can generate almost the same initial signal. Moreover, the noise generated by the operation of several small appliances or by the variation of electric voltage in the utility lines could mask the real signal. For this reason, in the second step, the hypothesis is tested, using additional features of each class of objects such as the variation of electric current in time or the frequency of activation. The hybrid combination strategy [10] is preferred to connect the rules used for the recognition of objects: (i) some rules are applied sequentially to reduce the preliminary set of possible objects and (ii) other rules are applied concurrently and independently, and their results are integrated using weighting factors.

The generic algorithms presented in this paper are developed from the analysis of electric current, monitored in a test house over a training period of 1 week, October 14–19, 1996. The equipment consists of clamp-on sensors and data loggers to record the measurements, performed every 16 s. This equipment is relatively inexpensive and can be easily installed, without power interruption. The energy signature of each appliance could very well be developed in terms of electric current. However, the electric demand is better perceived by the potential users or readers. Therefore, the energy signature is developed in terms of electric demand, using a constant voltage. This substitution does not affect the accuracy of results, mainly the evaluation of energy share, which is defined as the contribution of each appliance to the whole-house electricity consumption.

Fig. 1 shows, as an example, the variation of electric demand at the main entry (herein called 'total') over 1-h interval on October 16, 1996 in the test house, as well as the electric demand of major appliances (domestic hot water heater, stove, refrigerator and baseboard heater). From the training data, a sample of events corresponding to the selected appliances is selected. The response of the total demand profile to the activation of selected appliances is assessed. This information is referred to as the appliance's energy signature or profile, and subsequently is translated into pattern-recognition rules. The algorithms are only based on monitored data and do not need information about the occupants' lifestyle and usage of appliances. Once the generic algorithms have been developed, they can be applied to other houses. However, some parameters used by the generic algorithms will have to be modified, through the training process, to fit the electric characteristics of appliances in each house. The authors believe that the generic rules presented in this paper can be applied to all hot water heaters and refrigerators. However, additional data will be collected in other houses to test: (i) the generic character of the proposed rules; (ii) the accuracy of results under different combinations of appliances and pattern of usage; and (iii) the impact of seasonal variation of energy usage. Finally, the additional data will help refine the method.

In the development of these pattern-recognition algorithms, it was assumed that the simultaneous events do not occur (e.g., two appliances are activated or deactivated at the same moment). This implies that an end or start event is not masked by another end or start event.

The algorithms are validated for three separate data periods: (i) a 6-day period in October (14th to the 19th) 1996, (ii) a 6-day period in November (20th to the 25th) 1996 to test the near-to-date applicability of the algorithms and (iii) a 7-day period in January (6th to the 12th) 1997 to test the far-to-date applicability of the algorithms.

The total electricity consumption of the case study house is disaggregated into the following categories: (i) domestic hot water heater, (ii) refrigerator and (iii) others. Separate algorithms for the DHW and the refrigerator are

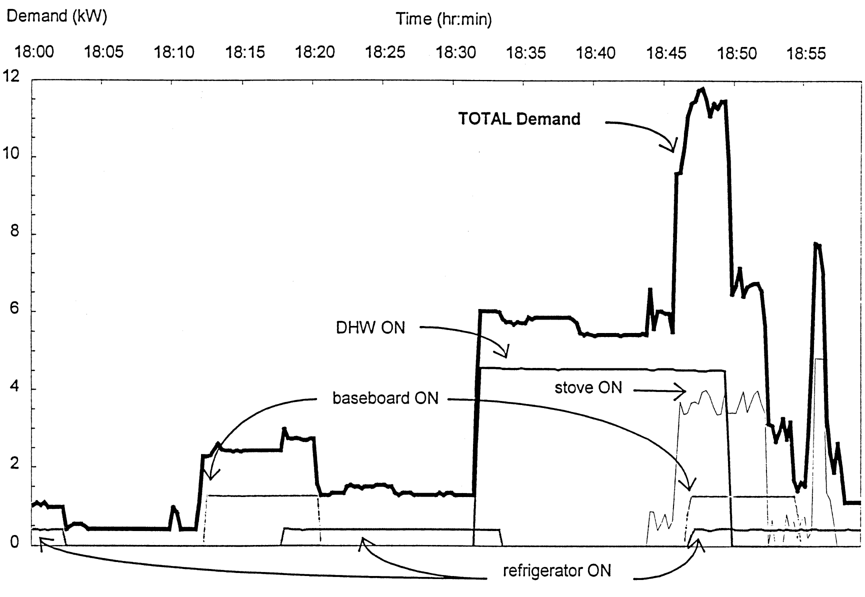

Fig. 1. Variation of the whole-house total electric demand and the electric demand of major appliances in the test house, during a 1-h interval.

developed. These two appliances are selected because they contribute to about 69% of the electricity used in the whole house. It is worth mentioning that the house has a hydronic heating system using oil as energy source. The two baseboard heaters are used as back-up sources only.

3. Algorithm for domestic hot water heater

The operation of the DHW heater is recognized from the total demand profile using a top-bottom rule-based algorithm. The algorithm consists of three stages (Fig. 2): (1) the detection of the activation (ON or start event) or deactivation (OFF or end event) of the appliance, by using its energy signature; (2) the estimation of the appliance's demand profile; and (3) the calculation of the appliance's energy use.

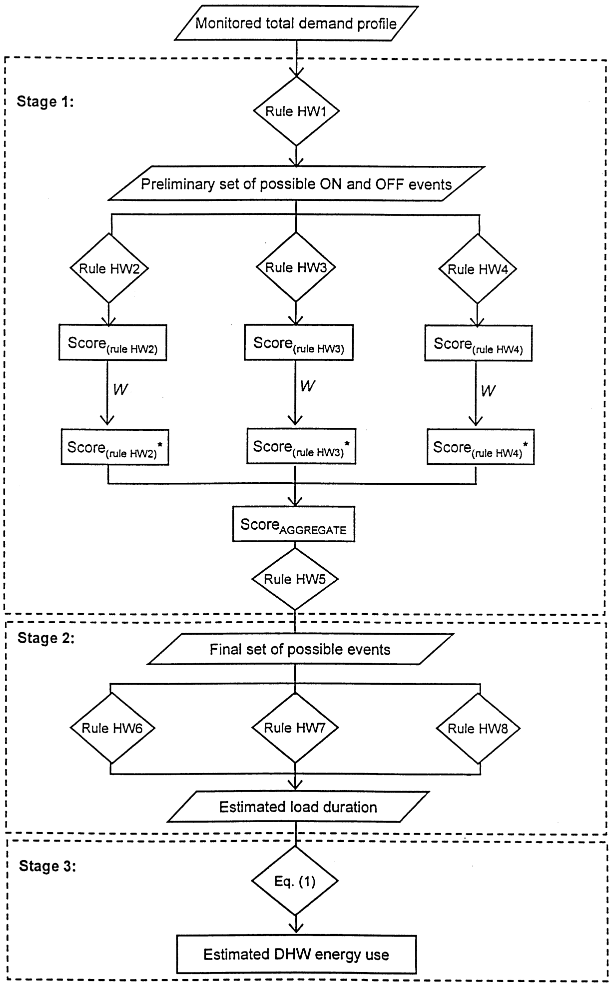

Fig. 2. Flowchart of the pattern recognition algorithm used to estimate the DHW energy use. Rule HW1—the state change detection rule. Rule HW2—the profile vector norm rule. Rule HW3—the number of data points rule. Rule HW4—the total demand rule. Rule HW5—the minimum score rule. Rule HW6—the highest scoring event rule. Rule HW7—the minimum ON and OFF interval rule. Rule HW8—the minimum total demand rule.

An appliance cycle is assumed to be composed of three separate segments: start, middle and end of each cycle (Fig. 3). The total demand is read on the left axis, whereas, the variation in the total demand between the time-steps (i) and $(i-1)$, represented as ΔTotal, is read on the right axis. The appliance energy signature is developed as a function of ΔTotal.

The boundaries for each segment are determined based on distinct changes in the demand pattern of the appliance load during a cycle. It was noticed that significant changes in the ΔTotal profile occur during the first six time-steps of a cycle, in the case of a start event, and in the last six time-steps of a cycle, in the case of an end event. The middle or steady-state segment remains approximately constant, therefore, this portion of the cycle cannot be easily recognized in the ΔTotal profile. Hence, the algorithm identifies the start and end events separately, and then joins the times of consecutive start and end events to obtain a sequence of intervals when the DHW is estimated to be in use. The duration of each of these estimated ON periods is then used to obtain the cumulative daily load duration of the DHW heater.

The first stage of the algorithm consists of five pattern-recognition rules, which are used to identify start and end events from the monitored total demand profile. The first rule, or the *state change detection* rule (HW1), is applied to determine a preliminary set of possible start and end events based solely on the detection of a given step-increase or decrease in the total demand profile. The following three rules, the *profile vector norm* rule (HW2), the *number of data points* rule (HW3) and the *total demand* rule (HW4), are then applied to each possible event. An aggregate score of the performance for these three rules is then attributed to each event. This aggregate score is then used to confirm or refute the occurrence of an event as recognized by the first rule. With the exception of the *number of data points* rule, the outcome of the scores of the *profile vector norm* rule and the *total demand* rule cannot directly refute an event. The *minimum score* rule (HW5) is used to filter out the weak events from the data set of possible events, that is, those events with low aggregate scores. The remaining set of selected start and end events proceed to the second stage of the algorithm, in which the daily load duration of the DWH heater is estimated.

The second stage of the algorithm consists of three constraining rules: the *highest scoring event* rule (HW6), the *minimum ON and OFF interval* rule (HW7) and the *minimum total demand* rule (HW8). Unlike the first stage of the algorithm, where the rules try to match the variation in total demand with the characteristic increase of each appliance, the goal of these rules is to verify that the DHW ON periods estimated in the first stage of the algorithm are consistent with the frequency of appliance usage, as obtained from the monitored data during the training period. Similar to the first rule, these rules are applied on a pass or

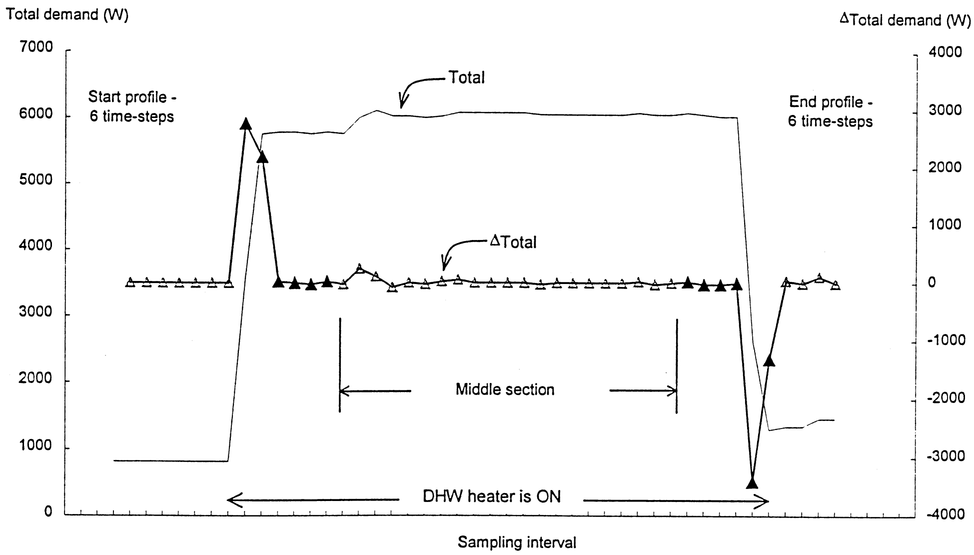

Fig. 3. Decomposition of the energy signature for the DHW heater.

fail basis. For example, an event is either confirmed as a start event or is recognized as a false start event, in which case, it is eliminated from the data set of possible start events. Consecutive start and end events are then linked together to obtain the sequence of activation periods of the DHW heater. Once the ON and OFF periods of the appliance are estimated, the DHW energy consumption for the duration of the day is then calculated as follows:

$$\text{Energy use} = \sum S_i \cdot \Delta t \cdot \text{Demand}_{\text{average}} \ [\text{Wh/day}]$$

where, S_i is 1 if the appliance is estimated to be ON at the time-step (i) or 0 if the appliance is estimated to be OFF, Δt is the sampling interval, equal to 16 s, and $\text{Demand}_{\text{average}}$ is the average electric demand of the appliance, which is obtained from the training data, and is equal to 4455 W.

3.1. Rule HW1: state change detection

The *state change detection* rule scans the ΔTotal demand profile for any step-increase or decrease that is within the range of variation of the DHW heater, monitored during the training period. This 'a priori' information is determined by analyzing the ΔTotal demand at the start and end of a sample of DHW cycles during the training period. The *state change detection* rule states:

$$\text{IF}\left(X_{\text{min-start-training}} \le x_i + x_{i+1} \le X_{\text{max-start-training}}\right)$$

THEN a possible start event occurs at time-step (i)

$$\text{IF}\left(X_{\text{min-end-training}} \ge x_n + x_{n-1} \ge X_{\text{max-end-training}}\right)$$

THEN a possible end event occurs at time-step (n)

where, x_i is the step-increase observed in the ΔTotal demand profile at time-step (i), and is calculated as $\text{Demand}_i - \text{Demand}_{i-1}$; $X_{\text{min-start-training}}$ is the minimum demand increase observed from the training data at the start of a cycle; $X_{\text{min-end-training}}$ is the minimum demand decrease observed from the training data at the end of a cycle; $X_{\text{max-start-training}}$ and $X_{\text{max-end-training}}$ are the artificial upper and lower limits imposed by the algorithm (9999 and −9999 W). Since the DHW heater has the highest demand of the appliances monitored, imposing upper limits on the step-increase or decrease is not necessary. However, this is a general algorithm that will be used to recognize other appliances as well. For example, for the refrigerator, actual monitored values for $X_{\text{max-start-training}}$ and $X_{\text{max-end-training}}$ are required to be input in order to differentiate the refrigerator's start and end events among those of other appliances. Hence, if the first part of rule HW1 yields, a start event is recognized at time-step (i). Similarly, if the second part of rule HW1 yields, an end event is recognized at time-step (n).

The step-increase considered by rule HW1 is the sum of x_i and x_{i+1}, instead of x_i alone. As well, the step-decrease is the sum of x_n and x_{n-1}, instead of x_n alone. The reason being that although, electric water heaters are steady-state appliances, depicted by discrete state-changes, the transient start or end events are sometimes measured, due to the small sampling interval, as being composed of two consecutive step-increases or decreases. The likelihood of this behavior appearing in monitored load data decreases as the sampling interval increases.

3.2. Rule HW2: profile vector norm

The *profile vector norm* rule consists of calculating the vector norm of a start or end demand profile with respect to the average start or end demand profile from the training period. This rule is based on the observation that the DHW heater start and end profiles are, in terms of pattern and demand level, distinguishable from those of other major appliances in the house. Fig. 4 shows some exam-

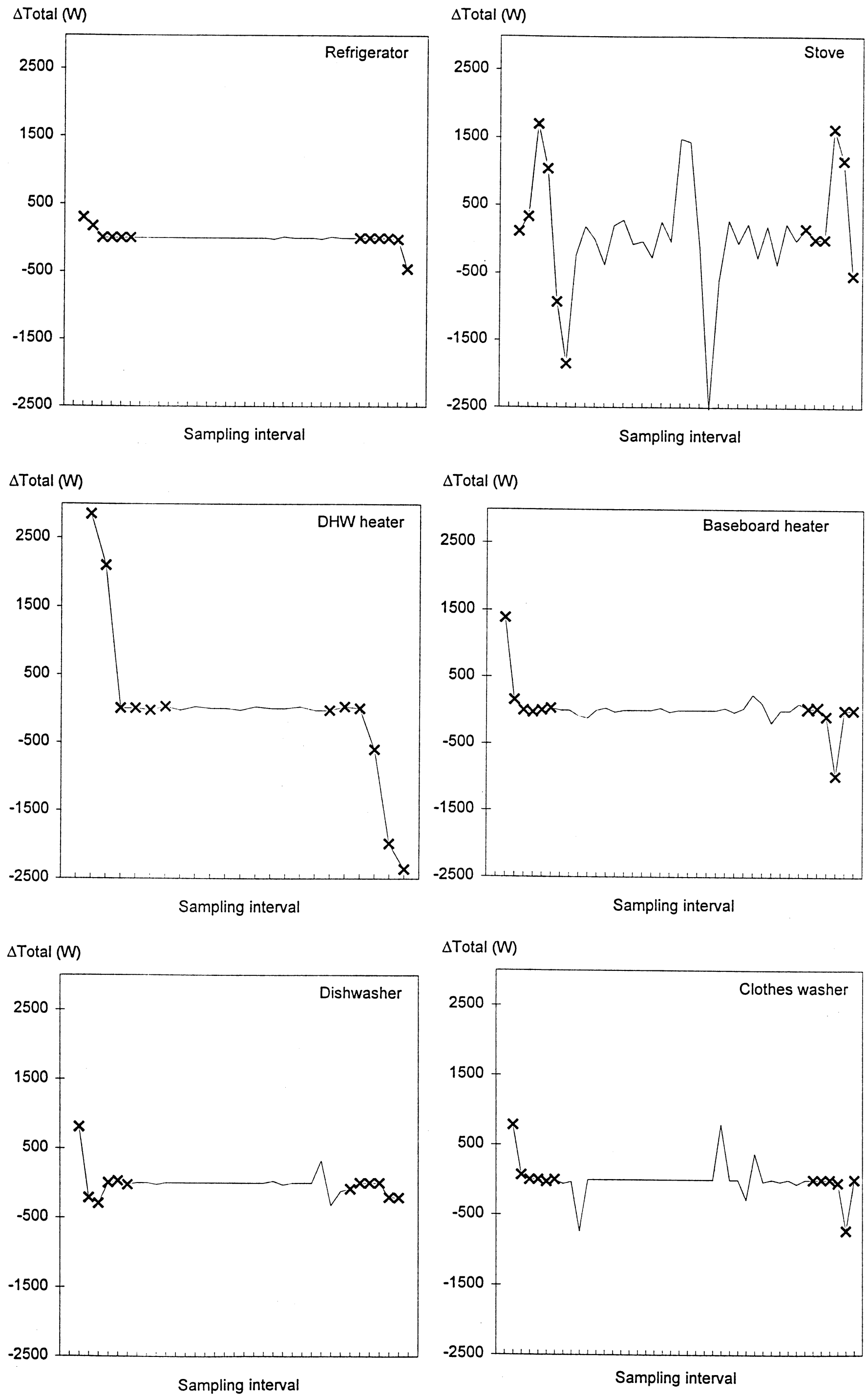

Fig. 4. Typical energy signature of major appliances.

ples of energy signatures of one cycle of operation, of the six major appliances monitored, as a function of ΔTotal. The highlighted data points illustrate the start and end profiles considered by the algorithm. The *profile vector norm* rule states:

$$\text{IF}\left(\text{norm}_i \leq \text{norm}_{\text{reference-start}}\right)$$

$$\text{THEN Score}_{\text{rule HW2}} = 1$$

$$\text{ELSE Score}_{\text{rule HW2}} = \text{norm}_{\text{reference-start}}/\text{norm}_i$$

$$\text{IF}\left(\text{norm}_n \leq \text{norm}_{\text{reference-end}}\right)$$

$$\text{THEN Score}_{\text{rule HW2}} = 1$$

$$\text{ELSE Score}_{\text{rule HW2}} = \text{norm}_{\text{reference-end}}/\text{norm}_n$$

where, $\text{norm}_{\text{reference-start}}$ is equal to 285 W (calculated as the average vector norm of 133 W plus two standard deviations of 76 W) and $\text{norm}_{\text{reference-end}}$ is equal to 776 W (calculated as the average vector norm of 564 W plus one standard deviation of 212 W), from a sample of demand profiles which occur during the training period. The vector norm values are calculated as follows:

$$\mathrm{norm}_i = \sqrt{\left\{\left[(x_i + x_{i+1}) - (C_i + C_{i+1})\right]^2 + (x_{i+2} - C_{i+2})^2 + (x_{i+3} - C_{i+3})^2 + (x_{i+4} - C_{i+4})^2 + (x_{i+5} - C_{i+5})^2\right\}}$$

where, x_i is the step-increase observed in the ΔTotal demand profile at time-step (i); C_i is the average start demand profile from the training period (Table 1). To calculate the vector norm for an end profile, the sequence of time-steps $(i,\ i+1,\ i+2,\ i+3,\ i+4$ and $i+5)$ is substituted by $(n,\ n-1,\ n-2,\ n-3,\ n-4$ and $n-5)$, where, (n) is the time-step identified as a possible end by rule HW1, and C_n is the average end demand profile from the training period.

3.3. Rule HW3: number of data points

The *number of data points* rule evaluates the number of time-steps in a possible start or end event whose ΔTotal demand falls between the upper and lower limits of variation of electric demand, established from the training data (Table 2). This rule is developed to reinforce those events that do not score well in rule HW2 because their profiles deviate from the expected demand profile. There must be some tolerance allowed between the expected profile and the measured profile to account for noise or variability in the total demand profile. The rule HW3 states:

IF ($X_{\text{min-start},i} \le x_i \le X_{\text{max-start},i}$ AND $X_{\text{min-start},i+1} \le x_{i+1} \le X_{\text{max-start},i+1}$
AND $X_{\text{min-start},i+2} \le x_{i+2} \le X_{\text{max-start},i+2}$)
THEN $\mathrm{Score}_{\text{rule HW3}} = 1$

Table 2
Upper and lower limits of variation of electric demand of the DHW heater [W]

Time-step	Start events		Time-step	End events	
	Upper	Lower		Lower	Upper
i	4920	1491	$n-2$	−27	27
$i+1$	3212	−54	$n-1$	−3370	27
$i+2$	27	−27	n	−4785	1034

ELSEIF (the above condition is only satisfied for two of the first three time-steps AND $\mathrm{norm}_i \le \mathrm{norm}_{\text{max-start}}$)
THEN $\mathrm{Score}_{\text{rule HW3}} = 0.5$
ELSE $\mathrm{Score}_{\text{rule HW3}} = 0$, and there is no possible start event at time-step (i).

To assess a possible end event using rule HW3, the sequence of time-steps $(i,\ i+1,\ i+2)$ is substituted by $(n,\ n-1,\ n-2)$, norm_i is substituted by norm_n, and $\mathrm{norm}_{\text{max-start}}$ is substituted by $\mathrm{norm}_{\text{max-end}}$.

3.4. Rule HW4: total demand

The *total demand* rule reinforces the recognition of the DHW heater when the total demand level is high. This rule is based on the assumption that the probability of recognizing the start or end event of the DHW heater increases as the demand of the whole-house increases. One can think that the total demand might be due to other end-uses, which operate concurrently and can add up to be equal to the demand level of the DHW heater alone. Therefore, an analysis of the total demand during DHW heater ON and OFF periods from the training period is done. The results show that there is a clear distinction between the two following scenarios: (i) a DHW heater in use alone and (ii)

Table 1
Demand profiles of the DHW heater, expressed as ΔTotal [W], for sample events of the training period

Time-step	Pure events						Mixed events				Average profile	Standard deviation
	1	2	3	4	5	6	1	2	3	4	C_i or C_n	
	Start events											
i	4757	4920	4704	1491	3910	3861	3678	2682	2620	2097	3472	
$i+1$	0	−54	135	3212	847	870	0	2154	3621	0	1079	
$i+2$	0	0	27	−27	0	0	−81	0	0	0	−8	
$i+3$	0	0	0	0	0	−27	−27	0	0	−27	−8	
$i+4$	0	0	0	27	0	27	0	−244	−766	27	−93	
$i+5$	0	0	0	0	0	0	0	0	−807	−27	−83	
Norm	75	113	104	71	75	76	438	115	839	1141	133	76
	End events											
$n-5$	27	54	−27	0	0	54	−27	0	−81	−27	−3	
$n-4$	−27	0	0	0	0	54	27	27	27	−27	8	
$n-3$	0	0	0	0	0	0	0	−27	−27	27	−3	
$n-2$	27	0	27	27	0	0	−27	0	0	0	5	
$n-1$	0	−2379	−1833	−874	−3266	−3370	27	0	−3169	−2748	−1761	
n	−4785	−2541	−3033	−3829	−1491	−1469	−4733	−4813	−1034	−2201	−2993	
Norm	788	286	60	398	673	721	800	788	676	450	564	212

a concurrent operation of several other appliances. The difference in the total demand for these two scenarios is about 2000 W. Hence, given the scenario that the DHW heater is ON and that the data is normally distributed, there is a very low chance ($\leq 2.5\%$) that the total demand of scenario (ii) exceeds that of scenario (i).

The scoring scheme for rule HW4 is evaluated using a continuous unipolar function, which has an asymptotic variation when the total demand approaches a very low and a very high level, that is, when the total demand is higher than two standard deviations above the mean, or is less than two standard deviations below the mean. The *total demand* rule states:

$$\mathrm{Score}_{\mathrm{rule\,HW4}} = \left[1 + \exp\left[-\lambda \cdot (\mathrm{TL}_i - \mu)\right]\right]^{-1}$$

where, the coefficient λ is 0.0017845, and is calculated by considering that 95% of all measurements of the total demand are within the mean plus or minus two standard deviations; the corresponding maximum total demand is 7432 W, and its score is 0.95 out of 1.0. The constant μ is equal to 5782 W, and is calculated as the average total demand observed during DHW heater ON periods from the training period; whereby, a score of 0.5 is attributed to the average total demand. The variable TL_i is the average total demand for time-step (i), calculated as the average demand at times $(i+2)$, $(i+3)$, $(i+4)$ and $(i+5)$. The demand at the time-steps (i) and $(i+1)$ is not used to calculate TL_i because the demand monitored during the first two time-steps is not representative of the steady-state demand level.

3.5. Rule HW5: minimum score

Before proceeding to the second stage of the algorithm, the preliminary data set of possible start and end events is evaluated using the *minimum score* rule and the aggregate score of rules HW2, HW3 and HW4. This rule is applied to remove false events which, most often, are characterized by low scores. The aggregate score is first normalized to a value between 0 and 1, using weighting factors to maximize the score of actual events and minimize the score of false events. This concept stems from the *generalized Delta rule* that is used to train a layered perceptron-type artificial neural network; according to the difference between the produced and target model outputs, the network's weights are adjusted to reduce the output error [11]. Similar to the *Delta rule*, the normalization weighting factors are determined based on the performance of actual start and end events observed during the training period. The average contribution of each rule's score to the aggregate score for actual events from the training period is then determined (Table 3). For instance, for the training day of October 14, the average contribution to the aggregate score, for all actual start events, is 0.35 for rule HW2, 0.38 for rule HW3 and 0.27 for rule HW4. Therefore, the selected start weighting factors, based on the entire training period, are the following: $W_{\mathrm{rule\,HW2}} = 0.40$, $W_{\mathrm{rule\,HW3}} = 0.40$ and $W_{\mathrm{rule\,HW4}} = 0.20$.

Once the aggregate score has been normalized, a minimum cut-off is applied to remove the false events from the data set of possible start and end events. Most often, these events have low aggregate scores. For the DHW heater the cut-off score is equal to the average aggregate score of actual events from the training period minus three standard deviations, that is, more than 95% of all events are included. This is equal to a score of 0.513 for the start events and 0.531 for the end events. Hence, the rounded value of 0.500 is applied for both start and end events.

Typically, actual start or end events from the training period yield fairly high scores, on average approximately 0.8 out of a maximum of 1.0 with an average standard deviation of about 0.1 for start events and 0.09 for end events.

The *minimum score* rule states:

$$\begin{aligned}
&\mathrm{IF}\ (\mathrm{Score}_{\mathrm{aggregate}\ i} > \mathrm{Score}_{\mathrm{cutoff\text{-}start}}),\ \text{whereby}\\
&\mathrm{Score}_{\mathrm{aggregate}\ i} = (\mathrm{Score}_{\mathrm{rule\,HW2}\,i} \cdot W_{\mathrm{rule\,HW2}})\\
&\qquad + (\mathrm{Score}_{\mathrm{rule\,HW3}\,i} \cdot W_{\mathrm{rule\,HW3}})\\
&\qquad + (\mathrm{Score}_{\mathrm{rule\,HW4}\,i} \cdot W_{\mathrm{rule\,HW4}})\\
&\text{THEN a start event occurs at time} - \text{step}\ (i)\\
&\text{ELSE there is no start event at time} - \text{step}\ (i)
\end{aligned}$$

Table 3
Selection of weighting factors used to normalize the aggregate score for the DHW heater

Training day of October 1996	Start events			End events		
	$W_{\mathrm{rule\,HW2}}$	$W_{\mathrm{rule\,HW3}}$	$W_{\mathrm{rule\,HW4}}$	$W_{\mathrm{rule\,HW2}}$	$W_{\mathrm{rule\,HW3}}$	$W_{\mathrm{rule\,HW4}}$
14	0.35	0.38	0.27	0.43	0.31	0.26
15	0.40	0.41	0.19	0.41	0.40	0.19
16	0.42	0.43	0.15	0.43	0.43	0.15
17	0.40	0.41	0.19	0.43	0.41	0.15
18	0.39	0.40	0.21	0.43	0.44	0.13
19	0.37	0.39	0.24	0.46	0.32	0.22
Average	0.39	0.40	0.21	0.43	0.39	0.18
Selected weighted factor	0.40	0.40	0.20	0.40	0.40	0.20

Where, $\text{Score}_{\text{cutoff-start}}$ is 0.5. The events recognized by this rule form the final set of start events, which will be used by the second stage of the algorithm. Similarly, rule HW5 applies to the end events, such that, $\text{Score}_{\text{aggregate } i}$ is substituted by $\text{Score}_{\text{aggregate } n}$, $\text{Score}_{\text{cutoff-start}}$ is substituted by $\text{Score}_{\text{cutoff-end}}$, and the time-step (i) is substituted by (n).

3.6. Rule HW6: highest scoring event

The *highest scoring event* rule verifies that the selected start event has the highest score among events within a range of 30 time-steps before (i) and 30 time-steps after (i). Similarly, the rule verifies that the end event selected has the highest score among events within a range of 30 time-steps before and after the event (n). The reason for imposing this constraint is again to remove false events that have gone undetected so far. The *highest scoring event* rule states:

$$\text{IF } S_i = 1 \text{ AND } \text{Score}_{\text{aggregate } i} > \text{Max } \{\text{Score}_{\text{aggregate } i-30}, \ldots, \text{Score}_{\text{aggregate } i}, \ldots, \text{Score}_{\text{aggregate } i+30}\}$$

$$\text{THEN } S_i = 1$$

$$\text{ELSE } S_i = 0$$

where, $S_i = 1$ indicates that the appliance is estimated to be ON and $S_i = 0$ indicates that the appliance is estimated to be OFF. Similarly, to select the highest scoring end event, the range $(i-30, \ldots, i+30)$ is substituted by the range $(n-30, \ldots, n+30)$; and S_i is replaced by S_n. The range $(i-30, \ldots, i+30)$ is selected based on the observation that the minimum duration of activation of the DHW heater, from the training period, is 19 time-steps (304 s), and the minimum time interval between two consecutive cycles is 13 time-steps (208 s).

Hence, the events for which $S_i = 0$ are eliminated from the list of potential events and those with $S_i = 1$ proceed to the next stage in the application of the algorithm.

3.7. Rule HW7: minimum on and off interval

The *minimum ON and OFF interval* rule verifies that a proposed change in state of the appliance (OFF to ON or ON to OFF) is in accordance with the usual pattern of usage of the appliance, as seen during the training period. In essence, the rule backtracks in the estimated DHW heater demand profile and confirms or refutes previous results. The rule states that a positive change in state of the appliance, that is, from OFF to ON at time-step (i) is likely to occur if the appliance was estimated to be OFF for a minimum of 10 sampling intervals prior to time-step (i). Similarly, a negative change in state of the appliance, that is, from ON to OFF at time-step (n) is likely to occur if the appliance was estimated to be ON for a minimum of 15 time-steps prior to time-step (n). The *minimum ON and OFF interval* rule states:

$$\text{IF } \left(S_i = 1 \text{ AND } \Sigma S_{i-j} = 0\right), \text{ for } j = 1 \text{ to } 10,$$

$$\text{THEN } S_i = 1, \text{ the start event occurs at time} - \text{step } (i)$$

$$\text{ELSE } S_i = 0, \text{ there is no start event at time} - \text{step } (i)$$

$$\text{ELSEIF } \left(S_n = 0 \text{ and } \Sigma S_{i-j} = 15\right), \text{ for } j = 1 \text{ to } 15,$$

$$\text{THEN } S_n = 0, \text{ the end event occurs at time} - \text{step } (n)$$

$$\text{ELSE } S_i = 1, \text{ there is no end event at time} - \text{step } (n)$$

$$\text{ELSE } (S_i = S_{i-1} \text{ AND } S_n = S_{n-1})$$

where, $S_i = 1$ or $S_n = 1$ indicates that the appliance is estimated to be ON, and $S_i = 0$ or $S_n = 0$ indicates that the appliance is estimated to be OFF.

3.8. Rule HW8: minimum total demand

The *minimum total demand* rule verifies that the total demand, measured during the activation of the DHW heater, is at least as high as the minimum observed steady-state demand for the DHW heater from the training period. The *minimum total demand* rule states:

$$\text{IF } \left(\Sigma S_{i-j} = 5 \text{ AND } \text{TL}_{i-2} < \text{TL}_{\text{min-total-DHW}}\right),$$

$$\text{for } j = 1 \text{ to } 5,$$

$$\text{THEN } S_i = 0$$

Table 4
Errors in estimating the daily energy consumption of the DHW heater

Day	Energy use [%]	Load duration [%]	Actual ON intervals missed [%]
Training period: October 1996			
14	−10.5	−9.8	13.8
15	0.3	0.5	0.0
16	−1.4	0.0	0.0
17	−0.2	−0.6	0.2
18	5.8	6.0	0.0
19	−0.7	−0.2	6.8
Average error	3.2	2.9	3.5
Near-to-date testing period: November 1996			
25	12.6	10.3	0.0
26	2.4	0.6	0.1
27	1.8	0.2	0.6
28	−3.7	−5.6	6.4
29	6.2	3.4	0.0
30	−3.0	−4.1	13.7
Average error	3.3	4.0	3.5
Far-to-date testing period: January 1997			
6	0.2	0.2	0.4
7	6.9	7.3	0.1
8	2.3	2.4	0.1
9	15.9	15.4	0.0
10	4.0	4.4	0.0
11	−7.1	−7.0	9.3
12	10.6	10.6	0.0
Average error	6.7	6.8	1.4

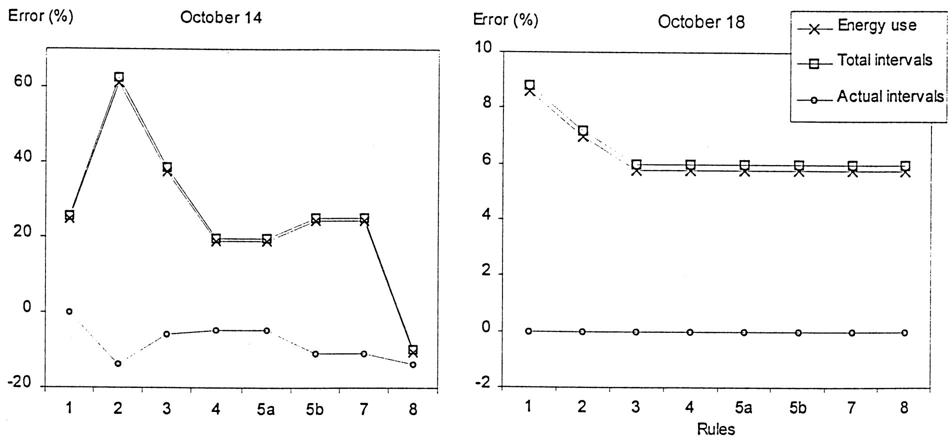

Fig. 5. Impact of each rule on the accuracy of the DHW heater algorithm for selected training days.

where, $TL_{min\text{-}total\text{-}DHW}$ is the minimum observed total demand level during operation of the DHW from the training period (2748 W). This value is less than the average steady-state demand monitored for the DHW heater (4455 W) because the first two time-steps of start and end profiles are considered.

4. Discussion of results for domestic hot water heater

The accuracy of the algorithm is assessed in terms of three performance indices: (i) daily DHW energy consumption, (ii) daily DHW demand profile and (iii) daily energy share, that is the ratio of the DHW energy consumption to the whole-house energy consumption.

4.1. Daily DHW energy consumption

The error in estimating the DHW energy consumption is calculated with respect to the monitored DHW energy consumption (Table 4). Negative values indicate that the algorithm underestimated the actual energy consumption, whereas, positive values indicate an overestimation. The average error is calculated using the absolute values of the

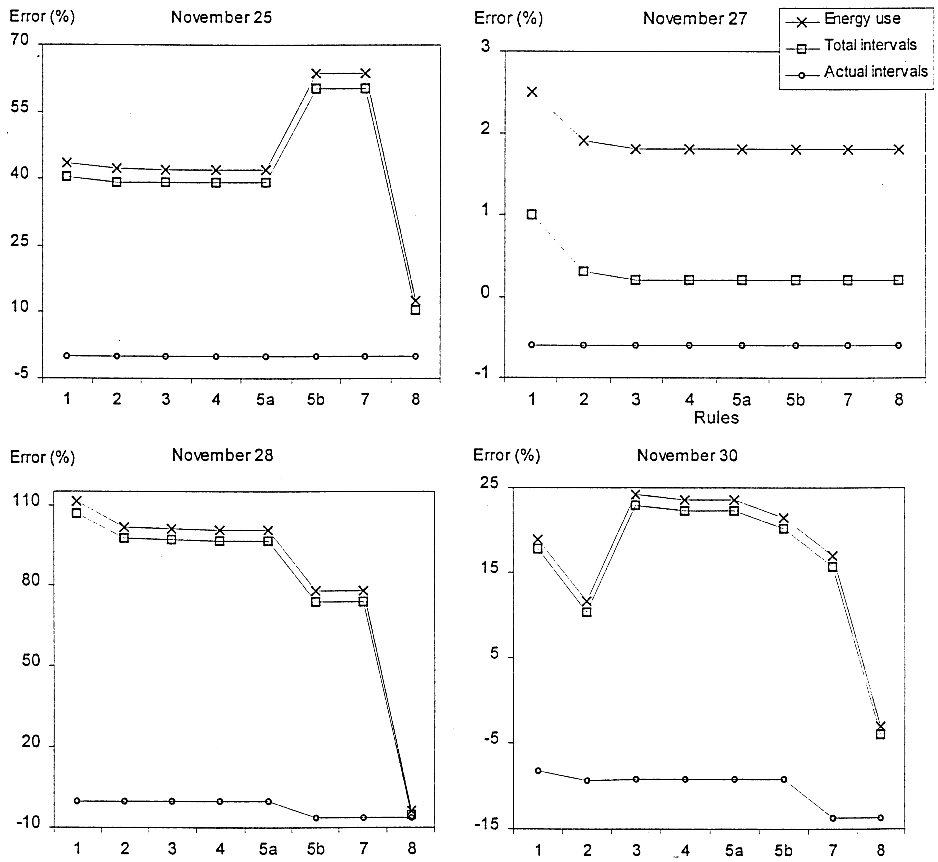

Fig. 6. Impact of each rule on the accuracy of the DHW heater algorithm for selected near-to-date testing days.

daily errors. This value is under 7% for all periods, that is, 3.2%, 3.3% and 6.7% for each period, respectively. At first glance, it appears that there is trend of higher error for the far-to-date testing period than the near-to-date testing period and the training period. This would indicate that there are seasonal differences in the appliance energy signatures, which the algorithm does not consider. However, on an individual day basis, the errors are fairly constant with the exception of 4 out of 19 days, that yield errors between −10.5% and 15.9%.

The estimation of the DHW load duration appears to have the largest impact on the error in estimating the daily energy use. The errors in estimating the load duration exceed 10% only for 3 days (November 25, and January 9 and 12).

The algorithm recognizes fairly well the time intervals when the DHW heater is ON. Only for 2 days, the percentage of actual intervals missed is 13.7% and 13.8%, respectively. This indicates that the algorithm can correctly recognize the actual start and end events. However, due to some false events, most often triggered by the stove, the algorithm overestimates the actual DHW heater energy use.

4.2. Daily DHW demand profile

Two statistics are used to evaluate the accuracy of the estimated DHW demand profile compared to the monitored profile: the coefficient of determination (R^2) and the average error, which is calculated using the absolute values of errors for each individual day. For any given day, the estimated profile tracks the monitored profile fairly well. The average error varies between 11 W and 312 W, that is, 0.2% and 7% of the average electric demand of 4455 W, observed during the training period. The coefficient of determination has values between 0.80 and 0.99.

4.3. Daily energy share of the DHW heater

From the homeowner's point of view, the most useful result is the estimation of the energy share, that is, the contribution of this appliance to the whole-house energy consumption and cost. For any given day, the difference in estimating the DHW heater energy share is between −5.1% and 5.5%. On a weekly basis, the average difference is consistently low, that is, 1.4%, 2.1% and 2.5% for the training period, near-to-date testing period, and far-to-date testing period, respectively.

4.4. Usefulness of the DHW heater rules

One can now question the need of eight rules in the pattern-recognition algorithm for the DHW heater. To respond to this question, a sensitivity analysis of the impact of each rule on the results of the DHW heater

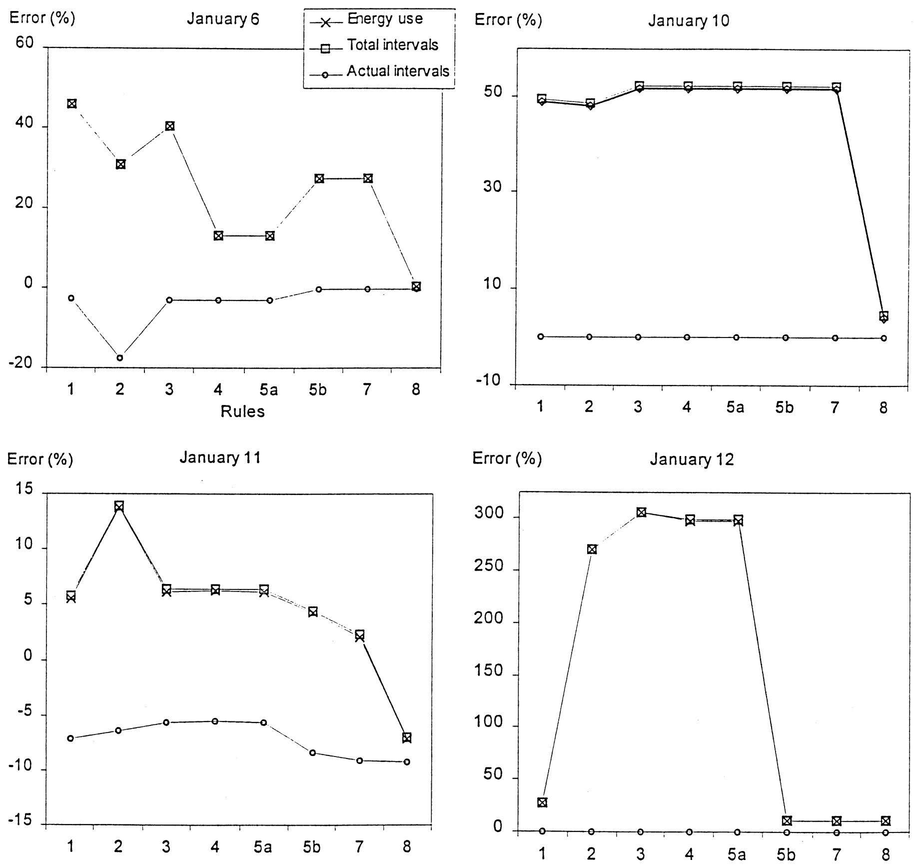

Fig. 7. Impact of each rule on the accuracy of the DHW heater algorithm for selected far-to-date testing days.

algorithm is performed. The sensitivity analysis focuses on the following three indices of accuracy: (i) daily DHW energy consumption, (ii) daily DHW demand profile and (iii) daily energy share of the DHW heater.

The analysis starts by applying the rule HW1 only to the entire data set, that is, assuming that a start or end event could occur every time there is a step-increase or decrease, which is within the recognized limits for the DHW heater (labelled Run 1). For the next runs (Run 2 to Run 8) a new rule is added to the previous set of rules, consecutively. Rule HW6, the *highest scoring event rule*, is applied in conjunction with rules HW2, HW3, HW4 and HW5. The following runs are performed:

Run 1: Rule HW1 only.
Run 2: Rules HW1, HW2, and HW6.
Run 3: Rules HW1, HW2, HW3, and HW6.
Run 4: Rules HW1, HW2, HW3, HW4, and HW6.
Run 5: Rules HW1, HW2, HW3, HW4, HW5a, and HW6.
Run 6: Rules HW1, HW2, HW3, HW4, HW5a, HW5b, and HW6.
Run 7: Rules HW1, HW2, HW3, HW4, HW5a, HW5b, HW6, and HW7.
Run 8: Rules HW1, HW2, HW3, HW4, HW5a, HW5b, HW6, HW7, and HW8.

Rule HW5, the *minimum score rule* is represented as two rules, the first part (HW5a) consists of applying the weighting factors, and the second part (HW5b) consists of applying the cut-off.

The results show the rules do not influence the accuracy of estimation in the same way, and their impact depends on the pattern of electricity usage in the house (Figs. 5–7). For instance, the addition of rule HW5b results in the increase of errors in estimating the energy use on November 25. However, the addition of the same rule HW5b results in a significant improvement of the accuracy for November 28 and 30 and January 11 and 12. Another example, is rule HW8 which on October 18 has no impact on the accuracy of the algorithm; however, it reduces significantly the errors on October 14, November 25 and 28, and January 10. Therefore, the combination of all eight rules is necessary to minimize the errors in the estimation of the DHW heater energy use.

5. Algorithm for refrigerator

The operation of the refrigerator is recognized from the total demand profile using a top–bottom rule-based algorithm (Fig. 8). Similar to the DHW heater algorithm, the refrigerator algorithm consists of three stages: (1) the detection of ON and OFF events, by using the energy signature, (2) the estimation of the appliance's demand profile, and (3) the calculation of the appliance's energy use.

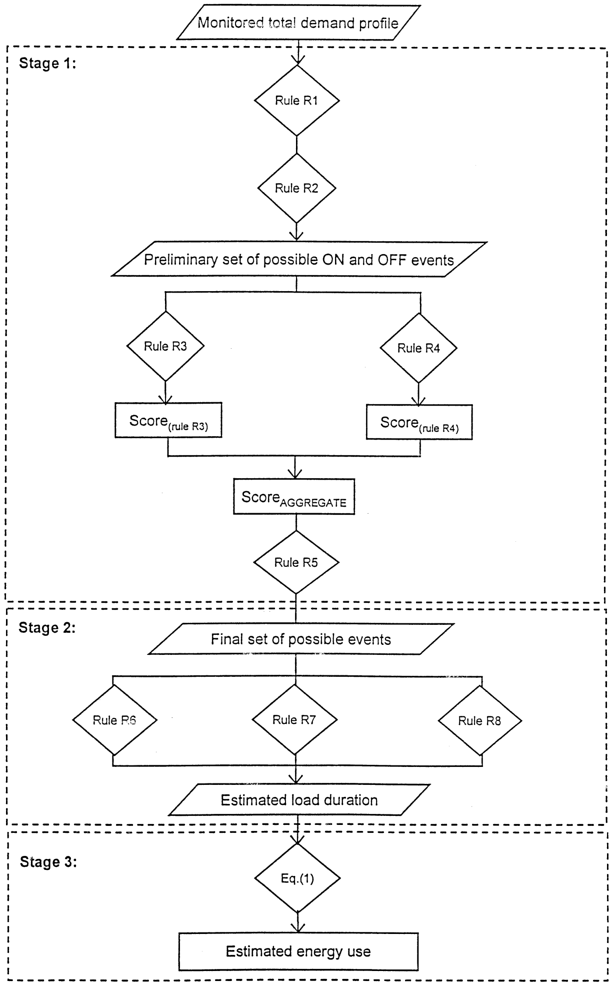

Fig. 8. Flowchart of the pattern recognition algorithm used to estimate the refrigerator energy use. Rule R1—the state change detection rule. Rule R2—the baseboard false event rule. Rule R3—the profile vector norm. Rule R4—the number of data points rule. Rule R5—the minimum score rule. Rule R6—the highest scoring event rule. Rule R7—the minimum ON and OFF interval rule. Rule R8—the minimum total demand rule.

The first stage of the refrigerator algorithm consists of five pattern-recognition rules, which are used to identify the start and end events based on the analysis of the monitored total demand profile. The first two rules are the *state change detection* rule (R1, similar to HW1 rule) and the *baseboard false event* rule (R2). The *baseboard false event* rule excludes those events that might be due to the start-up of the electric baseboard heater. It was found that the first or second time-step of a baseboard event could yield similar characteristics as that of the refrigerator. Therefore, rule R2 is applied to minimize the number of false events. Based on the application of R1 and R2, a preliminary set of possible start and end events is established.

Next, the following two rules: the *profile vector norm* (R3, similar to HW2 rule) and the *number of data points*

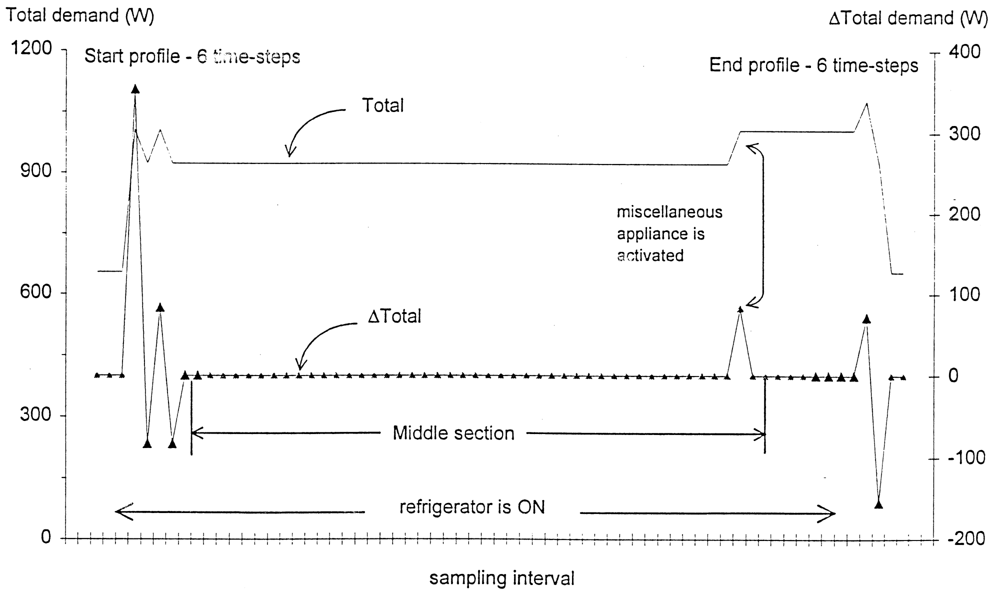

Fig. 9. Decomposition of the energy signature for the refrigerator.

(R4, similar to HW3 rule) are applied to each possible event. Next, the *minimum score* rule (R5, similar to HW5 rule) is used to calculate an aggregate score of performance of the events, and to eliminate the weak events from the data set of possible events.

The remaining set of selected start and end events proceed to the second stage of the algorithm, in which the daily load duration of the refrigerator is estimated. The second stage of the algorithm consists of applying three constraining rules: the *highest scoring event* rule (R6, similar to HW6), the *minimum ON and OFF interval* rule (R7, similar to HW7) and the *minimum total demand* rule (R8, similar to HW8).

Consecutive start and end events are then linked together to obtain the sequence of intervals during which the refrigerator is estimated to be in use. Once the ON and OFF periods of the appliance are estimated, the refrigerator energy consumption for the duration of the day is then calculated, using an average electric demand of 407 W, obtained from the training period.

Similar to the DHW heater algorithm, the refrigerator cycle is assumed to be composed of three separate segments: start, middle, and end of each cycle (Fig. 9). A spike at the beginning and end of each cycle distinguishes the refrigerator demand profile from that of other major appliances in the house. Similar to the DHW heater, it is found that significant changes in the ΔTotal profile are captured within the first six time-steps of a cycle, in the case of a start event, and in the last six time-steps of a cycle, in the case of an end event.

6. Discussion of results for the refrigerator

The accuracy of the algorithm is assessed in terms of three performance indices: (1) daily energy consumption, (2) daily demand profile and (3) daily energy share, that is the ratio of the refrigerator energy consumption to the whole-house energy consumption.

6.1. Daily refrigerator energy consumption

The average error in estimating the daily energy consumption is equal to 3.4%, 6.0%, and 7.7% for each

Table 5
Errors in estimating the daily energy consumption of the refrigerator

Day	Energy use [%]	Load duration [%]	Actual ON intervals missed [%]
Training period: October 1996			
14	−1.2	−0.5	12.9
15	0.2	1.8	6.6
16	−3.2	−1.0	8.8
17	5.4	6.9	3.3
18	0.6	2.0	7.5
19	9.5	11.9	5.8
Average error	3.4	4.0	7.5
Near-to-date testing period: November 1996			
25	3.9	4.0	7.6
26	3.0	2.6	10.7
27	4.3	3.2	9.6
28	8.9	7.6	13.0
29	13.9	12.1	10.5
30	1.9	2.0	10.7
Average error	6.0	5.3	10.4
Far-to-date testing period: January 1997			
6	−10.4	−9.3	18.5
7	−6.5	−5.2	17.8
8	11.6	13.1	10.6
9	−2.1	−0.5	10.5
10	−2.2	−1.3	11.5
11	15.1	16.9	11.0
12	−6.2	−4.4	14.6
Average error	7.7	7.2	13.5

period, respectively (Table 5). The error in estimating the daily load duration exceeds 10% only for 4 days.

6.2. Daily refrigerator demand profile

For any given day, the estimated profile follows fairly well the monitored profile. For each data period, the average coefficient of determination is 0.82, 0.73 and 0.72, respectively. The average error of the estimated profile varies between 18 W and 108 W, that is, from 4% to 27% of the average electric demand of 407 W, observed during the training period.

6.3. Daily energy share of the refrigerator

For any given day, the difference between the monitored and estimated energy share is between 1.7% and −1.2%. On a weekly basis, the average difference is 0.6%, 0.9%, and 0.8% for the training period, near-to-date testing period, and far-to-date testing period, respectively.

6.4. Usefulness of the refrigerator rules

Similar to the section dedicated to the DHW heater, a sensitivity analysis of the impact of each rule on the results of the refrigerator algorithm is performed.

The analysis starts by applying the rule R1 only to the entire data set, that is, assuming that a start or end event could occur at each time-step whose step-increase or decrease is within the recognized limits for the refrigerator (labelled Run 1). For the next runs (Run 2 to Run 7) a new rule is added to the previous set of rules, consecutively. The following runs are performed:

Run 1: Rule R1 only.
Run 2: Rules R1 and R2.
Run 3: Rules R1, R2, R3, and R6.
Run 4: Rules R1, R2, R3, R4, and R6.
Run 5: Rules R1, R2, R3, R4, R5a, and R6.
Run 6: Rules R1, R2, R3, R4, R5a, R6, and R7.
Run 7: Rules R1, R2, R3, R4, R5a, R6, R7, and R8.

A trade-off is noticed between (i) the estimation of the refrigerator energy use and (ii) the estimation of the number of actual refrigerator intervals which are recognized. For instance, for all days presented in Fig. 10, the rule R4 reduces the number of actual refrigerator intervals missed by the algorithm, but also increases the error in estimating the refrigerator energy use. Rules R7 and R8 significantly reduce the error in estimating the refrigerator energy use for all days. In most cases, the combination of all eight rules is necessary to minimize the errors in the estimation of the refrigerator energy use.

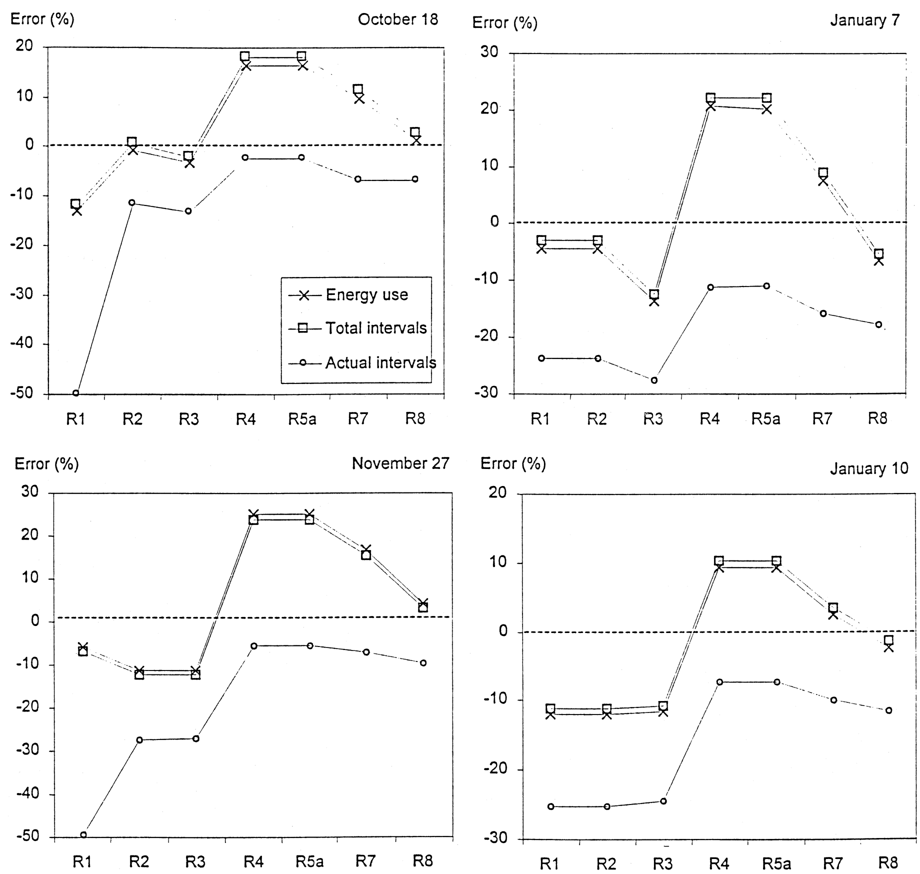

Fig. 10. Impact of each rule on the accuracy of the refrigerator algorithm for selected days.

7. Conclusions

The method presented in this paper shows a promising potential for application in residential buildings. The results prove that the whole-house electricity consumption can be disaggregated into its major end-uses, using a pattern recognition approach and only one sensor installed on the main electric entrance of the house. It also required a one-time submetering of the target appliances during the training period, of about a week, to find the electric characteristics of appliances.

The accuracy of estimating the contribution of each appliance to the whole-house electricity consumption and cost is fairly good. The proposed method was tested with monitored data from 3 weeks: (i) the training period of 1 week in October, (ii) the near-to-date testing period of 1 week in November, and (iii) the far-to-date testing period of 1 week in January. For instance, the utility bills of October 1996 of the test house indicate the total electricity consumption was 912 kW h at a cost of $60.6. Should the system presented in this paper would be installed, the difference between monitored and estimated contribution of the two major appliances is as follows: (i) 13 kW h or $0.85 for the DHW heater, and (ii) 6 kW h or $0.36 for the refrigerator. The overall difference for both appliances does not exceed $1.25 for the month of October, which appears to be acceptable for every homeowner. The errors in evaluating the daily energy consumption is between −10.5% and 15.9% for both the DWH heater and the refrigerator.

The authors believe that the generic rules presented in this paper can be applied to all hot water heaters and refrigerators. Moreover, the rules for a particular appliance should not be dependent on other appliances of the house, unless their electric characteristics are quite similar. In this case, additional rules should differentiate between the appliances. Additional data will be collected in other houses to test: (i) the generic character of the proposed rules, (ii) the accuracy of results under different combinations of appliances and pattern of usage, and (iii) the impact of seasonal variation of energy usage. Finally, the additional data will help to refine the method.

Acknowledgements

The authors acknowledge the support received from the Natural Science and Engineering Research Council of Canada and from the FCAR-Centre de recherche.

References

[1] G.W. Hart, Non-intrusive appliance load monitoring, Proc. IEEE 80 (12) (1992) 1870–1891.
[2] L.K. Norford, S.B. Leeb, Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load detection algorithms, Energy and Buildings, No. 24, 1996, pp. 51–64.
[3] T.R. Sharp, Non-intrusive load monitoring systems: considerations for use and potential applications, ACEEE Summer Study on Energy Efficiency in Buildings, No. 2, 1994, pp. 241–247.
[4] S. Yamagami, H. Nakamura, A. Meier, Non-intrusive submetering of residential gas appliances, ACEEE Summer Study on Energy Efficiency in Buildings 1 (1996) 265–273.
[5] J. Powers, B. Margossian, B. Smith, Using a rule-based algorithm to disaggregate end-use load profiles from premise-level data, IEEE Computer Applications in Power, April 1991, pp. 42–47.
[6] J. Powers, M. Martinez, End-use profiles from whole house data: a rule-based approach, ACEEE Summer Study on Energy Efficiency in Buildings 4 (1992) 193–199.
[7] B. Margossian, Deriving end-use load profiles without end-use metering: results of recent validation studies, ACEEE Summer Study on Energy Efficiency in Buildings 2 (1994) 217–223.
[8] Telephone conversations with Mr. Derrick Rebello of Quantum Consulting, 1998.
[9] P. Grey, Psychology, Chap. 9: Perception. Worth Publishers, 1997.
[10] A.F.R. Rahman, M.C. Fairhurst, An evaluation of multi-expert configurations for the recognition of handwritten numerals, Pattern Recognition 31 (9) (1998) 1255–1273.
[11] D.C. Park et al., Electric load forecasting using an artificial neural network, IEEE Trans. Power Systems 6 (2) (1991) .