End-Use Profiles from Whole House Data: A Rule-Based Approach

John T. Powers, Quantum Consulting, Inc.
Mark Martinez, Southern California Edison

Utility demand-side program evaluators have come to appreciate the value of end-use load data for accurate, defensible impact evaluations. However, the collection of metered end-use load data is often a time-consuming and expensive process, especially if the utility has not previously undertaken end-use metering projects. This paper describes an innovative method for developing accurate hourly residential end-use load shapes without relying on end-use metered data. We describe a rule-based disaggregation approach which makes full use of existing premise-level interval load data. Many utilities already gather such for rate design, system planning, or in response to the Public Utilities Regulatory Policies Act (PURPA). In the past, several statistical approaches have been considered to disaggregate premise-level load profiles, but have not overcome a number of analytical and applications challenges. The algorithm described in this paper has overcome these challenges, using common-sense rules for pattern recognition.

We first present a description of our rule-based load shape disaggregation algorithm. Next, we describe several validation studies, in which the algorithm has been tested at sites with both premise and end-use level load data. Finally, we give an overview of several demand-side program studies which have successfully made use of this disaggregation algorithm in both the design and evaluation phases. These applications include integrated end-use and premise-level sample design, evaluation of new end-use technologies, and impact assessment for load management strategies.

Introduction

Increasingly, utility planners and demand-side program evaluators are relying on the collection of end-use specific load and energy data. However, the collection of such end-use data is often a time-consuming and expensive process, especially if the utility has not previously undertaken end-use metering projects. This paper discusses a new technique that will allow utilities and other users to model accurately the effects of residential demand-side management programs using premise-level load data in place of end-use data. In the past, several statistical approaches have been considered to disaggregate premise-level load profiles, but have not overcome a number of analytical and applications challenges. The rule-based algorithm described in this paper has overcome these challenges.

The primary advantage of using premise-level (total household) load data to produce disaggregated end-use load profiles is cost. Many utilities have already collected such load data for rate design, system planning or in response to the Public Utility Regulatory Policies Act (PURPA). For establishing typical baseline end-use information, this whole-premise data is an excellent resource, since residential load research samples are generally designed to be statistically representative of the entire sector or rate class.

For other purposes such as the evaluation of specific demand-side programs, experimental designs which call for whole-premise metering rather than end-use metering can be implemented more quickly and at a small fraction of the cost. Depending on the technology used, end-use metering can be as much as five times more expensive than whole-premise metering. By planning the analysis techniques to be used, including end-use disaggregation, an experimental design for program evaluation can be far more cost effective and accurate than relying on end-use metering alone. Many recent metering studies (see, for example, [Schaper et al., Cason et al.]) for evaluation of DSM programs ranging from direct load control to residential new construction have made use of integrated experimental designs with a small number of end-use metering sites and a much larger number of whole-premise metering sites. This integrated approach allows for enough metered end-use load data to properly calibrate the disaggregation model, while leveraging the end-use data with much less expensive premise-level data from a broader cross-section of the population.

This paper describes a disaggregation algorithm developed and successfully applied to several utilities' premise-level load data. The algorithm is implemented through a PC-based proprietary software package. As described below, this algorithm has produced reliable, accurate end-use (appliance) load profiles from premise-level load data. As such, the algorithm has increased the value of existing premise-level load data and assisted utilities in evaluating end-use technologies and demand-side programs. In several applications where end-use load profiles were available for comparison, the algorithm produced "heuristic" profiles that closely approximated the actual end-use load profiles. The algorithm allows utilities to realize immediately many of the benefits of end-use load monitoring projects at a fraction of the cost and time. The algorithm uses only premise-level data and appliance connected loads to construct heuristic load profiles for any end-use with large connected load. It can be used by utilities from any geographic region or customer mix. Thus, accurate end-use information can be obtained with the algorithm using existing load research data, or using whole-premise load data gathered from a sample designed specifically for program evaluation. The algorithm provides this information in the form of end-use load profiles and appliance Unit Energy Consumption (UEC) values. Figure 1 illustrates a schematic representation of how the algorithm works.

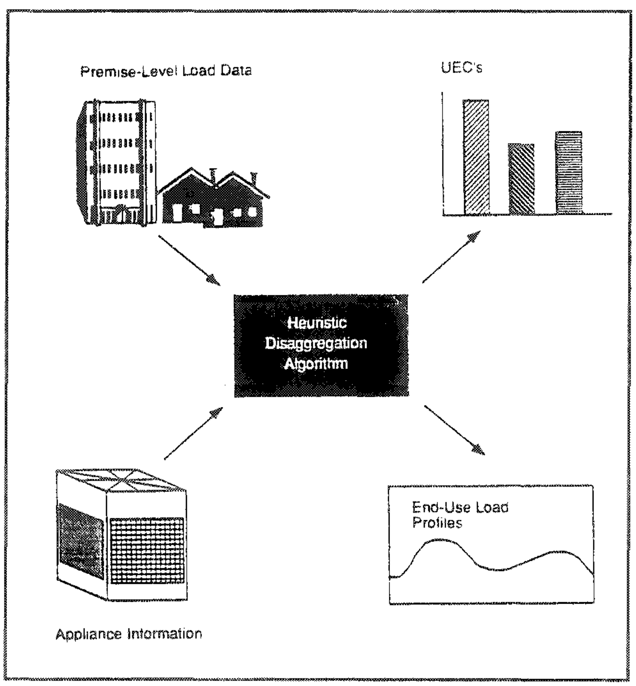

Figure 1. Schematic Representation of the Disaggregation Model

The algorithm draws its value from the fact that most utilities currently have important data that can support end-use analyses. Premise-level load research data have been collected by utilities from representative customer samples for years to satisfy PURPA or rate-making requirements. Appliance/equipment saturation data are collected from customers via detailed surveys on a periodic basis. On-site audit data often provide valuable engineering information about various end-uses. The algorithm ties all these types of data together to provide insights about customers' end-use profiles and usage patterns.

The Disaggregation Algorithm

The algorithm produces heuristic load profiles by applying a proprietary, rule-based disaggregation algorithm to premise-level load data. The algorithm uses information contained in each premise-day of load data. The algorithm combines these load data with information on end-use connected loads, customer appliance ownership data, and customer behavioral assumptions. These data are analyzed using a set of common-sense rules for pattern recognition. For a given premise-day, the algorithm scans the premise-level load profile and records the occurrence, timing, and magnitude of all large changes (or "spikes") in premise load. Next, the algorithm determines which of these spikes correspond to the end-use considered, and adjusts them according to consistency checks. The characteristics of each spike considered by the algorithm are magnitude, timing, and duration. These three characteristics, along with some information about previous and subsequent spikes at the same premise, are sufficient to determine whether or not that spike is attributable to the end use being analyzed. The load profile produced in this way is a disaggregated end-use load profile specific to the premise, appliance, and day being analyzed. In this way, the algorithm can be thought of as an "expert" looking at every single premise-day in the load research database and deducing the intervals over which the selected end use operated.

An example of how the disaggregation algorithm works is shown in Figure 2. There are two interesting observations of note in this figure, which shows the premise-level load profile along with its corresponding disaggregated central air conditioner profile. First, the only significant jump or spike in the premise-level profile for this date occurred at 11:00 am. The disaggregation algorithm classified this spike as being attributable to the central air conditioner because of its height (0.8 kWh or 3.2 kW) and duration (7 hours). Second, the disaggregation algorithm set the disaggregated central air conditioner profile equal to the premise-level profile minus the base load, calculated by inspecting the premise-level profile over periods when there is no significant energy usage. Figure 3 shows the premise-level and disaggregated central air conditioner profiles for a different premise-day. Note that the

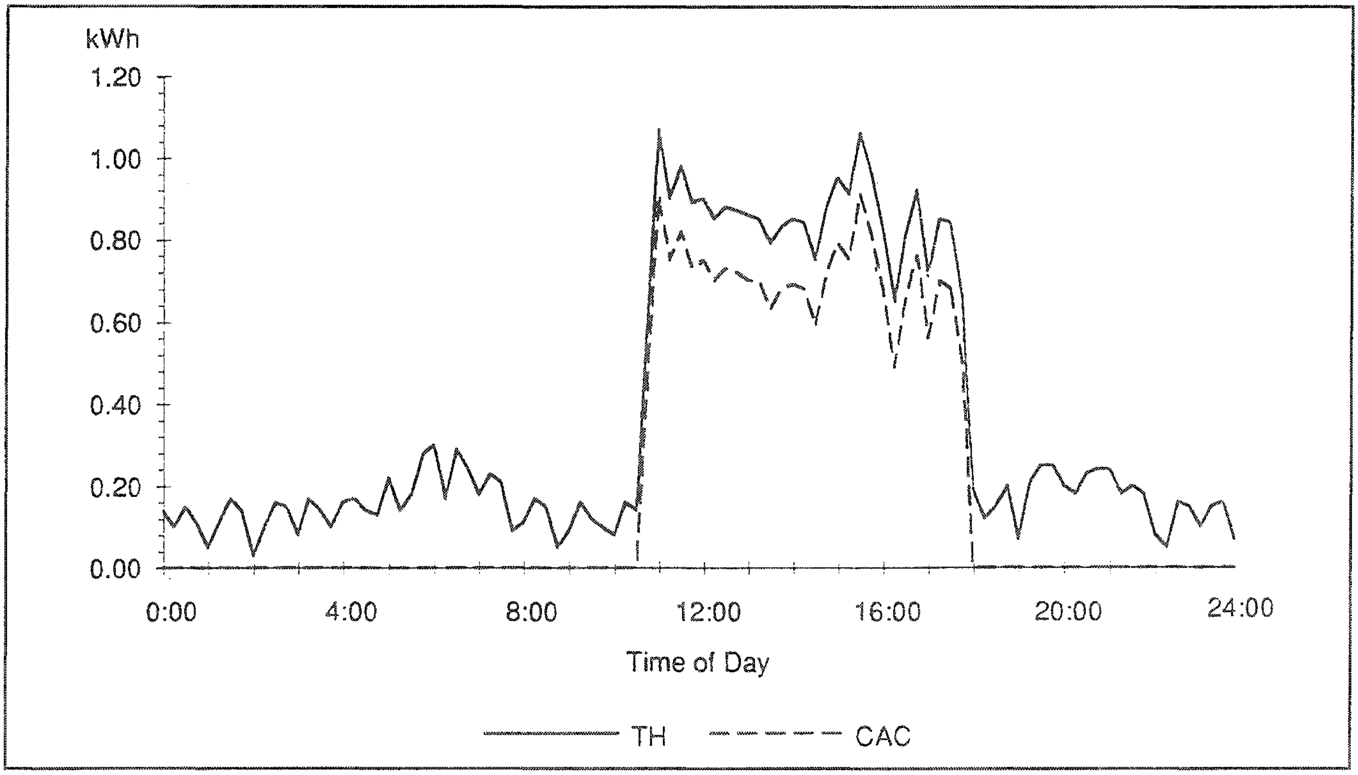

Figure 2. Disaggregation Algorithm: Example 1

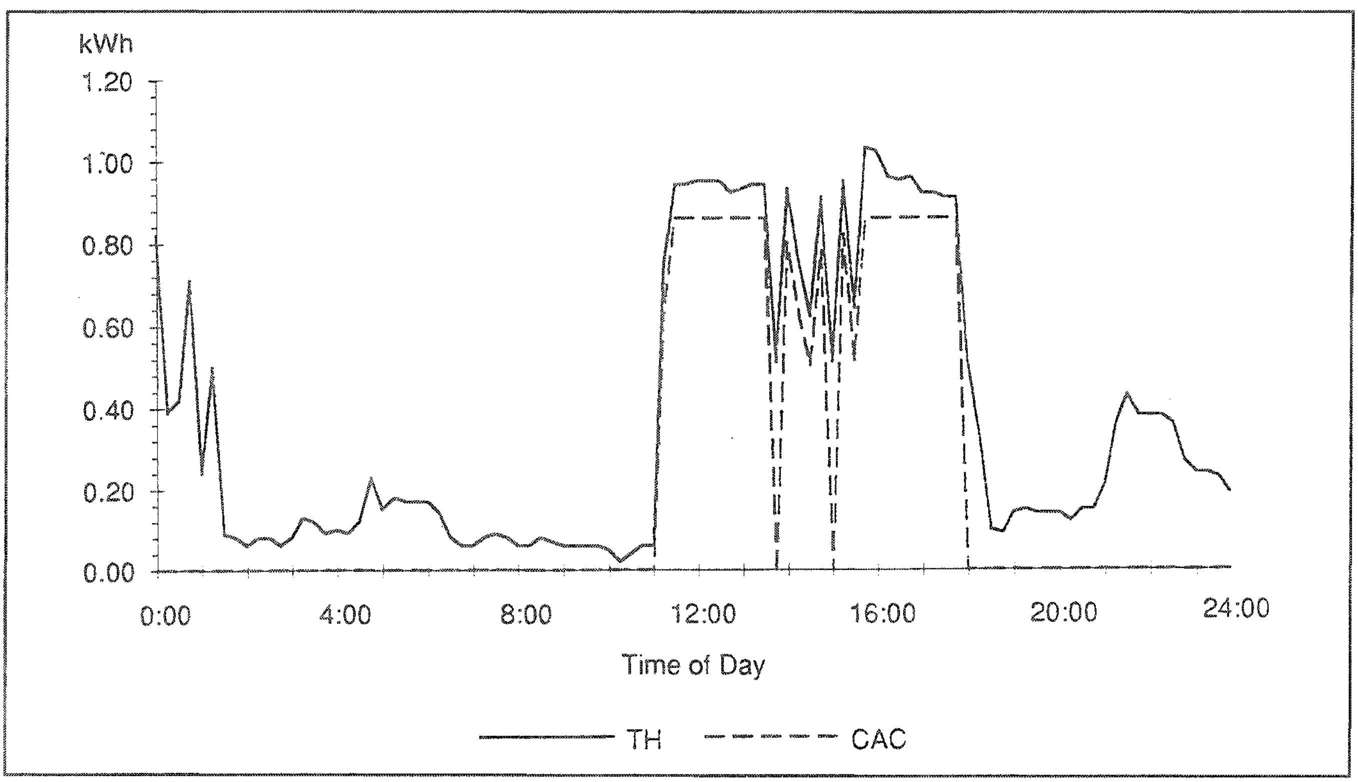

Figure 3. Disaggregation Algorithm: Example 2

disaggregated profile does not exceed 3.2 kW (or 0.8 kWh), as this is the connected load of the central air conditioner in this premise. The algorithm also ignored the spike that occurred around 9:00 pm since its height (around 0.8 kW) is too small relative to the connected load of the central air conditioner.

At first, it might appear that the algorithm as described above would be able to accurately derive the load profile for an end-use, if that end-use were the only major source of energy usage in the given premise. However, we have found that through a judicious, sequential use of this algorithm, accurate end-use profiles for all the major end-uses present in a premise can be generated. Whenever a premise is known to contain more than one major electric end-use appliance such as air conditioning and water heating, the algorithm proceeds to first disaggregate the largest end-use. The generated heuristic end-use profile is then subtracted from the premise-level profile. The algorithm then scans this "remainder" profile and proceeds to disaggregate the second largest end-use. This procedure is repeated until all major end-uses have been disaggregated from the premise-level profile.

As mentioned above, the algorithm relies on customer appliance/equipment ownership data as well as connected load estimates for all major end-uses that are to be disaggregated. The connected load estimates are used not only to match jumps in the premise-level load profile with the appropriate end-use, but also to allocate energy usage to the proper end-uses whenever two or more of the major end-uses are operating simultaneously. In cases where two or more of the major end-use appliances have similar connected loads, the algorithm distinguishes between them based on behavioral assumptions regarding the usage of these end-use appliances such as the time of day during which the usage occurs, the length of time over which the usage occurs, and the pattern of usage.

Other methods for disaggregating end-use profiles rely on sophisticated statistical techniques but fail to use all available data. Statistical disaggregation methods such as conditional demand analysis may produce good estimates of total energy usage for some end-uses, but have relatively little ability to accurately estimate load **shapes**. Further, because these techniques work by developing averages across premises, they are incapable of generating accurate profiles for **each individual** premise in the sample. The algorithm performs systematic, rule-based pattern recognition tests for appliance energy use over every individual measurement interval. Through this thorough use of all available information, the algorithm generates reliable load profiles for major end-uses.

Validation Studies

The algorithm has been successfully tested in several utility validation studies. These studies have been conducted for utilities in every region of the country, and for a wide variety of end-use technologies. The results of two of these studies demonstrate that the algorithm can produce reliable load profiles for residential appliances with large connected loads, such as air conditioners, heat pumps, and water heaters.

In each of these studies, both premise-level and end-use load data were available. This allowed us to compare the disaggregated load profiles (using only premise-level data) with actual end-use load profiles. In the first validation study, the algorithm produced reliable air conditioner load profiles. In another study, premise-level load data of a second utility were disaggregated to produce accurate heuristic water heater load profiles. In each study, the algorithm produced heuristic end-use load profiles tracked the corresponding metered load profiles very closely. The results for both studies are summarized below.

Summary of Disaggregation Results

The results of our validation studies show that the heuristic end-use load profiles correspond very closely to those calculated from the actual metered end-use load data. The peak value of the disaggregated air conditioner load profiles, when averaged over all households for all summer days, differ from the peak of the averaged metered profile by less than 5 percent. The average summer air conditioner Unit Energy Consumption (UEC) estimate derived from the heuristic load profiles differs from the actual UEC by less than 10 percent. The timing of the average disaggregated air conditioner peak is also predicted very accurately.

In the air conditioner study, the algorithm used premise-level and air conditioner load data from almost 40 households during four summer months. The resulting heuristic load profiles were averaged across premises and summer days and assessed by comparing them to the actual air conditioner load profiles. A similar procedure was used to derive heuristic load profiles in the water heater study.

Assessment of Heuristic Load Profiles

There are several ways that we have assessed the validity of the disaggregation algorithm. These ways include visually comparing the heuristic end-use load profiles with real load profiles as well as rigorous statistical tests. Described below are the results of our assessment of the algorithm for estimating end-use load profiles.

Load Profile Comparison. A straightforward way of assessing whether the heuristic load profiles produced accurately represent metered end-use load profiles is to make graphical comparisons, which we have done. A comparison of the average heuristic and average actual (metered) load profiles is shown in Figure 4. This figure shows that, on average, the heuristic air conditioner load profiles track the metered ones very closely. Both the timing and the size of the average heuristic air conditioner peak are consistent with the actual peak. The distance between the two profiles, on average, is only about 50 Watts. Figure 5 shows a similar comparison for the water heater study. As the figure shows, even with a much smaller sample, the average heuristic water heater load profile approximates the metered profile over most of the day. The heuristic profile captures the twin peaks that water heater profiles typically have. While the peaks occur at the same time in both profiles, the heuristic profile slightly underestimates the actual peak in this case.

Figure 6 shows a load profile comparison for a single premise-day for the air conditioner study. This figure shows a single customer's metered and heuristic air conditioner load profiles for one summer day, along with the premise-level load profile from which the heuristic profile was derived. As the figure illustrates, even for an individual premise-day, the heuristic load profile tracks the metered end-use profile throughout the day.

Two observations in Figure 6 are worth discussing. First, the algorithm correctly ignored the increases in the premise-level load profile at 8:00 a.m. and at noon. The algorithm correctly identified these spikes as usage of other appliances (not the central air conditioner) and therefore set the heuristic air conditioner load equal to zero over this time interval. Second, at 2:00 p.m., the algorithm recognized the increase in premise-level load as being caused by air conditioner usage.

The premise-day illustrated in Figure 6 is a typical example of how well the algorithm performed. While the algorithm worked well on the majority of premise-days considered, some premise-days were too noisy to accurately disaggregate.

Statistical Comparison. Another method for assessing the overall validity of the algorithm is to compute several statistics that summarize how consistent the heuristic load profiles are relative to the actual profiles. The summary statistics considered are listed below, together with a brief discussion of each.

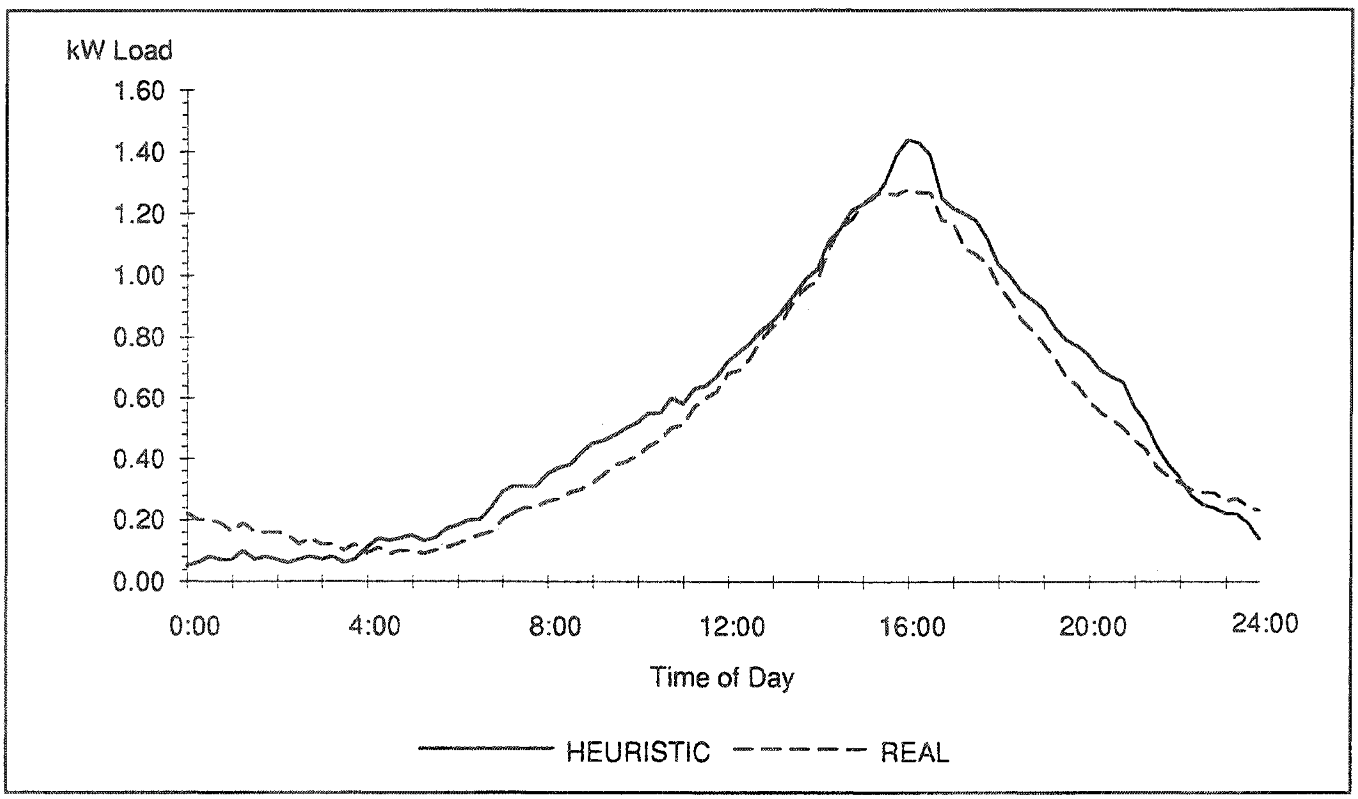

Figure 4. Comparison of Heuristic and Actual Air Conditioner Average Load Profiles

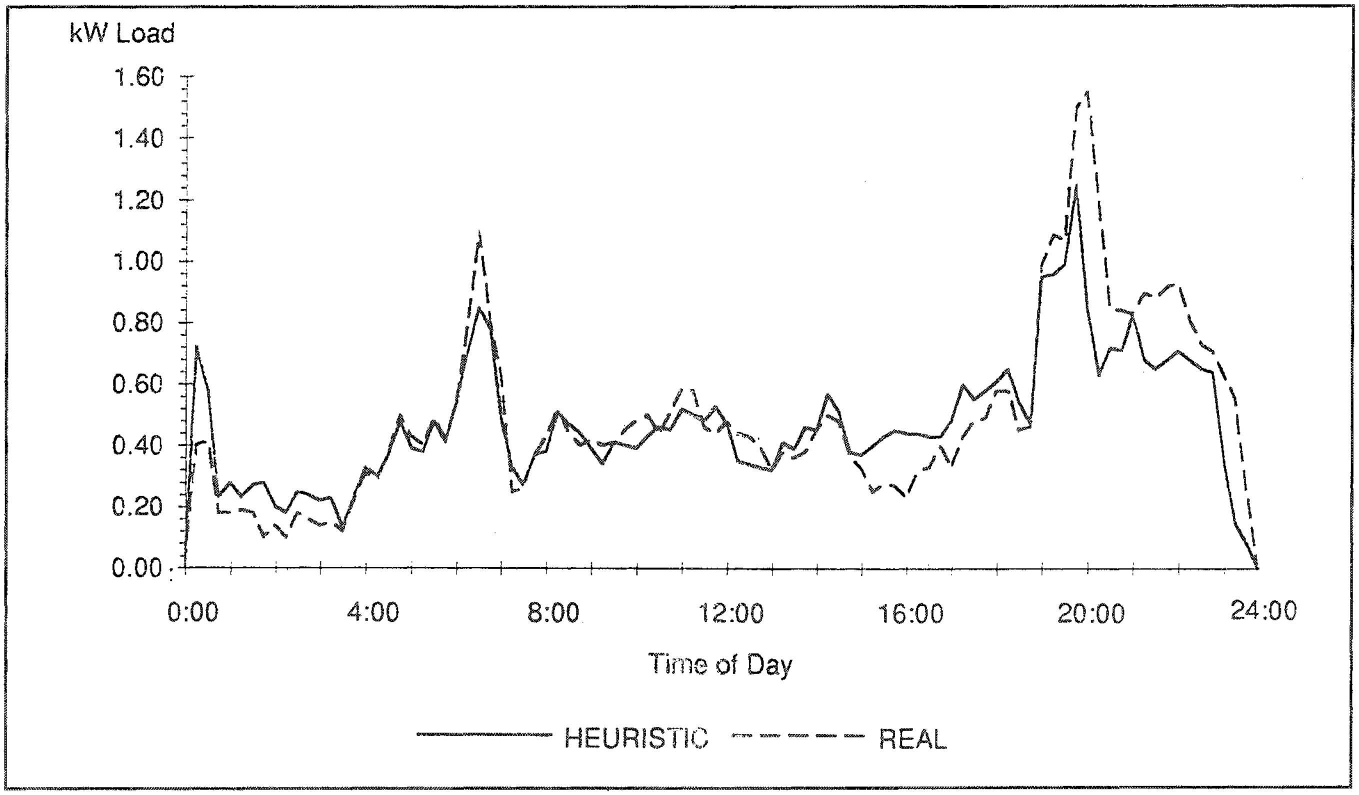

Figure 5. Comparison of Heuristic and Actual Water Heater Average Load Profiles

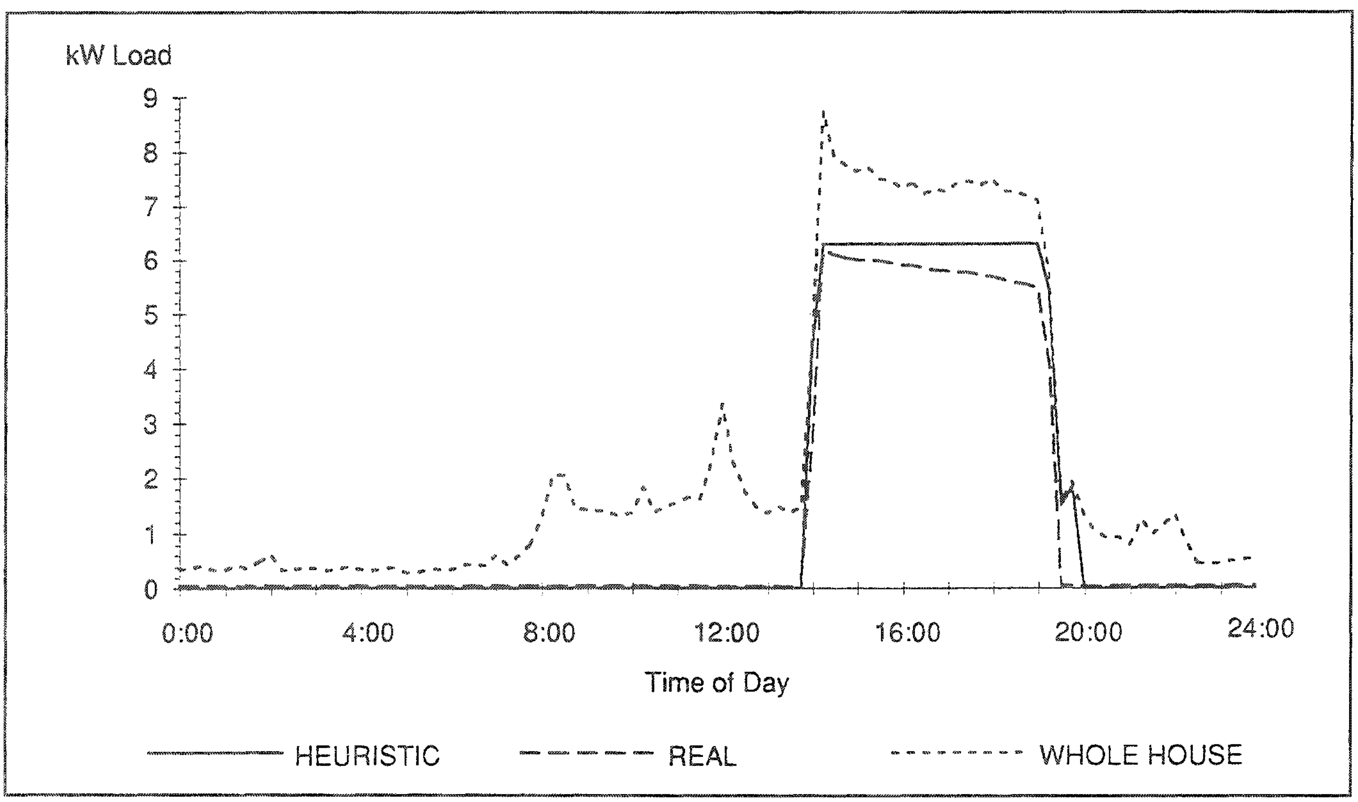

Figure 6. Heuristic and Actual Air Conditioner Load Profiles for One Premise-Day

Mean Squared Error: This statistic is computed by averaging the squares of the differences between the average actual and heuristic load profiles at each of the 96 observations in Figure 4. For the air conditioner study, the value of this statistic obtained was 0.004, that is, the average squared error between the average real and heuristic profiles is quite small.

R-Square: This statistic is computed by first evaluating the ratio of the sum of squared errors to the sum of the squared values of the real profile, and then by subtracting this result from one. The closer the value of this statistic is to one, the better is the fit between the heuristic and real profiles. For the air conditioner study, the value we obtained for this statistic was 0.989.

Mean Absolute Error: This statistic is computed by averaging the sum of the absolute value of the differences between the real and heuristic profiles. The resulting value is then the average kW difference between the two profiles. For the air conditioner study, the value we obtained for this statistic was 0.051 kW, or 51 Watts. The average connected load of the air conditioners being monitored was about 5400 Watts.

UEC Error: This is the difference between the real unit energy consumption (UEC) and the UEC implied by the heuristic profile. For the air conditioner study, the UEC error was 32 kWh, based on a 30-day month. In this study, the algorithm estimated the real air conditioners' summer UEC within 10 percent.

Future Directions

At present, the disaggregation algorithm is implemented using a decision tree. Because this can be a very labor-intensive approach, we also explored the possibility of using a neural network to implement the algorithm. This research was quite discouraging, given the promise of neural networks for pattern recognition. The patterns in premise-level load data which are apparent to load research experts are frequently not recognized by any neural network implementation we have attempted. This research continues, because we are convinced that there are substantial benefits in terms of further automating the disaggregation process which could be obtained once some technical difficulties are overcome.

Additionally, we are in the process of modifying the algorithm to include explicit consideration of weather conditions. This change should improve the accuracy of the heuristic end-use profiles, and may expand the list of appliances which can be successfully disaggregated.

Acknowledgments

This paper was prepared with the support of Southern California Edison Co. (SCE). The opinions, findings, conclusions and recommendations expressed herein are solely those of the authors, and do not necessarily reflect the view of SCE or Quantum Consulting. The authors wish to acknowledge the assistance of Mr. Bedig Margossian, Dr. Bruce Smith, and Ms. Nerissa de Jesus in preparing this paper.

References

Powers, John, et al. *Disaggregating Air Conditioner Load Shapes from Whole House Data — A Study.* April 1990.

Powers, John, et al. "Using a Rule-Based Algorithm to Disaggregate End-Use Load Profiles from Premise-Level Data." *IEEE Computer Applications in Power.* April 1991.

Schaper, Carl and Bruce Smith. *Residential Air Conditioner Direct Load Control Evaluation Plan.* March 1992.

Uhlaner, Robert and Carl Schaper. *Wisconsin Electric Direct Load Control Evaluation Plan.* December 1991.